

What is Europe's best way forward?

The path to European unity has been strewn with obstacles since the collapse of the European Exchange Rate Mechanism and will hardly be clarified at the summit meeting arranged for 29 October. But something can be done.

WHATEVER happened to the dream of a United States of Europe? These are dog days for the federalists. The Maastricht Treaty on which European governments have invested time they could not afford and, in some cases, have traded compromise for political influence to win the right to ratify the treaty, has been made something of a dead letter by the collapse of the Exchange Rate Mechanism at the beginning of August last; although nothing of principle has been overtly changed, so that the special meeting of the heads of European governments, known as the European Council, called for 29 October will almost certainly agree that the next stage towards monetary union (the first step towards creating a European central bank) should begin on 1 January, nobody is now talking of a timetable for a single currency. Matters may be further confused this week, when the German Constitutional Court rules on a complaint that the Maastricht Treaty robs German voters of their democratic rights to determine their own affairs. In other words, the European Communities (EC) will duly become the European Union (EU) on 1 November this year, but will not clearly know where they go from there.

Part of the difficulty is long-standing. Now, as always, the EC is in the habit of setting deadlines for the attainment of certain pre-conditions for the further development of its institutions. Thus Maastricht requires not only that there should be a single currency by 1999 at the latest, but that, before then, inflation rates and government deficits should be reduced below targets now considered to be generally unattainable. At a meeting of finance ministers and central bankers at the weekend, it seems to have been agreed that the best that can be done, for the months immediately ahead, is to do whatever possible to improve the chances of economic growth in Europe. Then, with luck, it should be possible to devise another timetable for monetary union, and the project for unifying Europe should be under way again.

The defect of that optimistic view is that it may also be unrealistic. M. Jacques Delors, the President of the European Commission (the EC's executive) is said to have warned last weekend's meeting that recession will still overhang Europe in 1996. Even that may be too hopeful, reckoning as Delors does without the claims on Europe's economies that are certain to mount in the years ahead. And while most members of the EC have made a start on making employment cheaper and state industries more competitive, the chances are only small that by 1996 the then EU will have made a serious start on rolling back the Common Agricultural Policy

which is now the most obvious brake on the economic development of Europe, but whose abolition would seriously injure many rural economies (throwing extra people out of work).

That is why the European project needs to be more guileful. While the chance is only small that the enrichment of European populations that has sustained the EC's development in the past four decades will continue in the years ahead, the troubles of the immediate past have also undermined many previously solid beliefs in the virtues of the EU. Britain may have ratified Maastricht, but only at the cost of fracturing political support for European goals. Even Germany, traditionally the pillar of European unity is now hesitant about the project of a unified Europe, ostensibly because the EC does not help with the immediate problems of unification, but perhaps also because unification is itself a substitute for the links with the outside world hitherto provided by the EC. But whatever hesitations there may be about the EU, nobody supposes that the omelette can be unscrambled at this stage.

That is why the meeting at the end of this month should be concerned to seek ways other than blind alleys in which to find success of some kind. Cooperation on foreign policy (required by Maastricht) is unlikely to be a fruitful field, given European disunity on ex-Yugoslavia. Similarly, the Maastricht edict that political union is a European goal will yield no quick benefits. But what the EC could usefully accomplish, and much more quickly, is to put a constructive stamp on social institutions from which the whole of Europe benefits. Education is one obvious field, research another. Maastricht will for the first time give the EC the right to a policy on education; it should set out to help make sense of European university systems. And the need to reshape the EC research programme in favour of basic research is long overdue. These issues should be on the agenda for 29 October. □

Crime and punishment

The British government's plans for law and order are potentially calamitous.

THAT criminality is an indicator of social incivility is nowhere disputed. It cannot be. For any crime is a sign that its perpetrators, at least, do not accept the rules by which the other members of a society have agreed that their conduct



should be constrained. Yet by this test, social incivility is rife. All countries are encumbered with criminality — and are perplexed at knowing how to deal with it. Governments deploy remedies of three kinds. Inevitably, there are penalties for convicted criminals, meant to deter them and their emulators or, in extreme cases, to prevent recidivism by imprisonment. The more efficient detection of criminals is another laudable goal; if the detection of certain crimes were virtually certain, those inclined to commit them would be powerfully deterred from doing so, while their victims could be equitably compensated. The third group of remedies, attempts to deal with the “causes” of criminality, range from the regulation of the instruments of some crimes (as with gun laws, for example) to attempts to moderate the alienation of which all criminality is a mark.

For decades now, it has everywhere been plain that certain conditions must be satisfied if any mixture of these remedies is even partially to succeed. First, the definition of criminality must command respect; when that is not the case (as during the prohibition of alcohol in the United States during the 1920s, possibly the use of cannabis now and the laws that deprive people of civil rights in every dictatorship), the definition of criminality is hopelessly blurred. Second, the administration of the system must be recognisably fair. If, for example, people are falsely accused and then falsely imprisoned or, alternatively, if alleged criminals escape penalties or even trial on insubstantial grounds, the respect that the law commands is inevitably undermined. All that is uncontentious.

Or, more accurately, might be thought to be so. But not, it seems, in Britain. In the past few years, several disconcerting happenings have cast doubt on the working of the criminal judicial system. In particular, there has been a string of convictions of people accused of terrorism which have been found on later examination to be flawed (in one case, after some of those convicted had spent sixteen years in jail). One common thread was found to be the over-sharp interpretation of forensic data by forensic scientists. Another was the concealment of relevant favourable evidence by the police and even, it has been alleged, the concoction of unfavourable evidence. Public anxiety at this apparent erosion of the judicial system reached such a point that the British government formed a Royal Commission under a judge, Lord Runciman, which reported only a few weeks ago (see *Nature* 364, 178; 1993). Yet last week, the trial of three ex-policemen charged (among other things) with having falsified evidence against falsely accused and jailed terrorists was abandoned by the trial judge on the grounds that the jury would have been prejudiced by all the publicity there has been.

That, in itself, would be bad enough. It is worse that these events coincided with the annual conference of the Conservative Party (which has formed the British Government for the past 14 years) and with what seems to have been the present government’s determination to “get tough with criminals”, as the Home Secretary, Michael Howard, put it last week. To that end, he announced a string of plans which, if carried through, may have an effect opposite to that

intended. The scheme to supplement fingerprinting by DNA “fingerprinting” for those accused of serious crimes (see page 595) is the least exceptionable of them; if careful consideration shows that police may find that an aid to their efficiency, there can be nothing against it.

Sadly, that is not true of the other decision, announced last week, that people who are accused of crimes, and who are released on bail pending trial, will be sent “straight to jail” if they are accused a second time. Howard did not tell his audience how that decision can be squared with the presumption of innocence until found guilty that has been taken as a cornerstone of British justice. Nor did he attempt to justify his decision that an arrested person’s exercise of the “right to silence” when questioned by police should not be used in a later trial as evidence for the prosecution, except to say that the right has been “ruthlessly exploited by terrorists” and that “the innocent have nothing to hide”. There are two difficulties. First, after a string of convictions acknowledged to be unjust and based on false confessions submitted as evidence by police, can the innocent be sure that what they say will not harm them? Worse, the Runciman Commission concluded that the present law is necessary if public respect for criminal justice is to be preserved. Howard’s decision that this recommendation will be over-ridden is hardly calculated to reinforce that respect.

The extenuating circumstances are familiar. Not the least is the setting at which Howard said that, in the past 30 years, “the balance in the Criminal Justice system has been tilted too far in favour of the criminal”; Conservative Party conferences are notoriously avid for “getting tough on criminals”. But where do those who unjustly spent sixteen years in jail fit in? Does Howard believe they were guilty after all? Or is it merely that he is prepared to contemplate a degree of rough justice for the sake of adding to the population of Britain’s already overcrowded jails? In either case, how does Howard calculate that his package of populist measures against those charged (but not yet convicted) of serious crimes will foster the partnership between the police and the public for which he pleaded at the outset of his lamentable speech?

It is, of course, a great misfortune that Britain, which has for many years prided itself on its system of justice, should abandon principle for prejudice (in the literal sense) in the manner now decreed. (To be fair, Scotland, with a legal system of its own, will be spared some of the planned indignities.) It is especially unfortunate that much of what has been learned of the working of judicial systems everywhere should have been so lightly set aside. Is it not, for example, worthwhile organizing imprisonment (which, Howard said last week, “works”) so that those imprisoned are less (rather than more) likely to emerge as confirmed criminals? And why not substitute the plan that the exercise of the right to silence may be used as evidence for the prosecution in a trial with a requirement that those arrested on suspicion of criminality should never be interviewed by the police except in the presence of independent persons? The trouble with Howard’s package is that while it will fill the jails, it will further undermine the credibility of the judicial system as a whole. □

India seeks to learn the lessons of the Maharashtra earthquake

New Delhi. The Indian Government has set up a panel of experts to investigate possible lapses by key scientific agencies responsible for monitoring and predicting earthquakes in the wake of the disastrous Maharashtra earthquake two weeks ago.

In particular, the panel — which has been given a month to report its conclusions — will look into allegations that vital seismic data about the region were not properly interpreted to assess the potential hazard. It will also investigate claims that senior geologists were quarrelling among themselves while strains were building up under Maharashtra.

At least 14 major agencies are involved in earthquake studies in India. P. R. Kumaramangalam, the science minister, says that the inquiry panel will not carry out a witch-hunt. But it “will examine all available seismic data pertaining to the region and assess if follow-up actions were taken by the agencies concerned”.

The panel will also suggest fresh surveys and studies of the region, for which the Ministry of Science and Technology is already seeking a \$25 million grant from the government. “We have been focusing on Himalayan quakes for too long”, says Kumaramangalam. “The time has come to look at the Deccan plateau, which we had assumed to be stable.”

The earthquake, which took place on 30 September, registered 6.4 on the Richter scale. It left roads and engineered constructions intact, but flattened houses made of mud and stone in 35 villages, killing about 12,000 people officially, although unoffi-



The Maharashtra earthquake occurred in a region 200 km west of Hyderabad, with low relative seismicity.

cial estimates are as high as 30,000. Scientists admit that the toll would have been lower if the state administration had been correctly warned about the hazard potential, and had pushed for the construction of earthquake-proof houses. According to Kumaramangalam, one of the jobs facing the panel is to find out whether the agencies had sufficient data to warrant an alert, and if so, why they failed to issue a warning.

The Maharashtra quake was the second major earthquake in living memory to hit the Indian peninsula, long considered to be seismically stable. The first, with a magnitude of 6.7, struck the Koyna reservoir in

1967. But most geologists, including Harsh Gupta, director of the National Geophysical Research Institute (NGRI) in Hyderabad, believed that it was caused by the weight of water in the reservoir.

Smaller quakes had been recorded in the Deccan plateau since then. But seismologists did not think it necessary to revise India's seismic zoning map. Indeed, the epicentre of the new earthquake lies in a zone at the bottom of a seismic scale of I to V (see figure). “It was a total surprise”, said Gupta.

But the earthquake has already exposed fissures in the scientific establishment. Janardhan G. Negi, until recently deputy director of NGRI, points out that he and his colleague, N. Krishna Brahmam, produced evidence in 1973 of a 390-km long rift valley under the Deccan traps. “We warned two decades ago that areas over this Kurdwadi rift needed careful monitoring, but nobody listened to us”, says Negi. “The epicentre of the present earthquake sits right on the rift.”

Negi, who is now science adviser to the Madhya Pradesh government, says that the epicentres of three earlier mild earthquakes also fell over the same rift, and points out that the Geological Survey of India recorded 125 tremors in the three months preceding the latest quake.

“These data, plus our prediction of a rift zone under this region, should have prompted NGRI to rush a few mobile monitoring stations to the site, but this did not happen”, says Negi, who claims that leading geologists failed to pay sufficient attention to what was happening in the region.

Gupta says the Maharashtra earthquake has forced a review of the seismic status of the Deccan plateau, but is not yet prepared to admit that the Kurdwadi fault exists. Gupta and H. N. Srivatsava, deputy chief of the Indian Meteorological Department (IMD), say that its existence must be confirmed by field observation.

A team from IMD, as well as another from Germany, are already at the site measuring micro-earthquakes. Srivatsava says that these data, when combined with records of 30 September earthquake from seismic observatories around the world, will allow scientists to locate the fault and understand the mechanism behind the quake.

Meanwhile the political fallout from the earthquake is already causing concern for the ruling parties both in Delhi and in the state of Maharashtra. One month away from elections in four states, opposition parties have launched a campaign criticizing ruling politicians for ignoring pleas for relocation from villagers who died. **K. S. Jayaraman**

David Dickson

Industrial take-over for UK science policy

London. Six of the ten non-government members of Britain's Council for Science and Technology, the newly created body responsible for providing advice to the government on issues of “strategic national importance”, are to be industrialists.

Announcing the members of the council on Monday, Prime Minister John Major said that these would include Richard Sykes, chief executive of Glaxo Holdings plc, Alan Rudge, director of development at British Telecom, and Sir Robin Nicholson, formerly science adviser to Margaret Thatcher and now director of Pilkington plc.

Nicholson was previously chairman of the Advisory Council on Science and Technology, the body whose role has been taken over by the new council following publication of the government's white paper on the organization of science earlier this year.

The other members of the committee —

which will be chaired by the minister for science, William Waldegrave — include Bridget Ogilvie, director of the Wellcome Trust, Sir Aaron Klug, director of the MRC Laboratory of Molecular Biology, and Stewart Sutherland, chancellor of the University of London.

Further evidence of the new thrust in science policy comes with the appointment of a top industrialist, widely credited for salvaging the British government's LINK scheme for bringing together university and industry research groups, as chairman of the Natural Environment Research Council.

In the first of a series of appointments flowing from the reorganization of British science outlined in the government's white paper, Robert Malpas, chairman of the Cookson Group plc, has been appointed to the new part-time position.

US funding agencies told to expect a 'new mission'

Washington. Eight months into the Clinton administration, a significant shift in emphasis from "science" to "technology" is gaining unexpected momentum. In its most recent manifestation, the chairman of an influential Senate committee has warned US scientists to expect a major reorientation of their work towards applied research.

"You are getting a new assignment", Senator John Rockefeller (Democrat, West Virginia), chairman of the Senate's Science, Technology and Space subcommittee, told the annual meeting of the National Academy of Engineering last week. "You should not be surprised to see Congress and the Clinton administration question the function of funding agencies", he said.

Rockefeller, whose committee has gained prestige from the ascension of its former chairman, Al Gore, to the vice presidency, spoke in particular of the need for change at the National Science Foundation (NSF).

He effectively backed strong language in the Senate's appropriations bill for NSF, which called on the agency to change its mission in the direction of so-called "strategic" research. The message for NSF in the bill, he said, was to "be more relevant to competitiveness, or don't expect more increases in your budget".

In an unusually blunt exposition which almost certainly reflects the administration's thinking on science policy, Rockefeller said

that the change would extend to other agencies, including NASA and even the National Institutes of Health. He said the need for change was shaped by three phenomena — the end of the Cold War, concern about US international competitiveness, and the views of Clinton and Gore.

In the past, Rockefeller said, winning the Cold War had been the country's top priority, and economic competitiveness had taken a back seat. Alone amongst industrial nations, the US government had made no investment in industrial technologies. The consequence, he said, was that the US found itself with "a major competitiveness problem".

Rockefeller warned that "the old era is over" for scientific research. He added that, despite the widespread impression to the contrary among university academics, this did not mean that Congress was criticizing science as such. But he conceded that many universities interpreted the report language on NSF as "an effort to convert them into applied research centres".

He said the aim was a more limited one, of telling the agencies to do more basic research in strategic areas, as defined by the Federal Coordinating Council for Science, Engineering and Technology. "Fifty-five per cent of NSF money already goes there. We would like that increased somewhat."

Colin Macilwain

Senior German health officials sacked

Munich. The president and director of Germany's federal public health bureau were sacked at short notice last week, leading to speculation that government officials had been suppressing information about the extent of HIV contamination of blood products in the 1980s.

But rather than exposing an AIDS scandal, the episode has drawn attention to the ineptitude of the government's top agency for overseeing health regulations.

Horst Seehofer, the health minister, had been complaining for months that the bureau, which is responsible for monitoring activities in areas such as pharmaceuticals, food and the environment, was neither running efficiently nor keeping him adequately informed on important issues.

Seehofer's patience finally ran out last Tuesday, when the bureau was unable to answer questions about a list of 373 cases of suspected HIV infections contracted through blood or blood products. The list had been put together through the bureau's department for drug registration, and had been presented following a meeting organized to

brief the minister on HIV risks.

The list itself revealed nothing new about HIV infection. Another list of 2,000 similarly infected patients, collated from reports from AIDS diagnosis laboratories, has long been in the public domain. Nearly all of these, including 364 of the partial list shown to Seehofer last week, were infected before the antibody testing and heat-inactivation of blood products became mandatory in 1985.

But the political ineptitude of the public health bureau in presenting the list at a time when the minister's temper was frayed is being seen as an illustration of the health office's apparent inability to make good judgements.

Seehofer complains that the bureau, which since reunification has grown to an unwieldy 3,800 staff scattered over many offices, mostly in Berlin, is unable to predict or prioritize the information he requires.

The precipitous sackings led to press speculation of an AIDS scandal involving a cover-up of the extent of infection due to careless handling of blood products.

Allison Abbott

NASA cites Concorde in boosting funds for aeronautics

Washington. The National Aeronautics and Space Administration (NASA) is to take the lead role in an enhanced federal research programme aimed at helping to restore the dominant position of the US civil aerospace industry in world markets.

NASA managers are increasing their emphasis on the aeronautics aspect of the agency's work, often forgotten by those who see it purely as a space agency. Congress has already agreed to sharply increase NASA's budget for aeronautics research and development from \$855 million last year to \$1,013 million in the financial year which begins this month.

NASA wants to raise the budget further. According to its administrator Dan Goldin, aeronautics will grow by 50 per cent by 1998; "and I don't think that's enough", he adds. The increase will be made at a time when NASA's budget for space research is being sharply curtailed.

John Dailey, NASA's acting deputy administrator, is currently leading a study of US aeronautics research facilities inside and outside of the agency, which will be completed early in the new year. NASA is likely to use the results of the study to press for funding to build new facilities, such as wind tunnels, for it to use in conjunction with leading aerospace firms such as Boeing, McDonnell Douglas and Pratt & Whitney.

Speaking last week at the National Academy of Engineering in Washington, Goldin claimed that America "had forgotten about aeronautics" since it gave up on building a supersonic airliner in the late 1960s. Goldin listed many faults of the Anglo-French supersonic plane, Concorde. But he said that its development programme had paved the way for the subsequent success of Airbus in winning 30 per cent of the world market in long-haul airliners from scratch.

NASA wants the support of the Clinton administration for a fully-fledged programme to develop an economically competitive supersonic airliner, but environmental as well as fiscal concerns make this unlikely for the time being.

The agency is therefore concentrating on attracting funds to boost existing long-term research programmes, and to build some of the aeronautics research facilities which the US aerospace industry desires. Goldin says that Pratt & Whitney and Boeing are reduced to using wind tunnels in France and the Netherlands for some of their work.

Most of the extra money will be spent at NASA's four aeronautics research centres: Langley at Hampton, Virginia; Lewis at Cleveland, Ohio; Ames at Mountain View, California and Dryden at Edwards, California.

Colin Macilwain

Watson to take the helm at Soros institute

Moscow. Nobel laureate James Watson is to take over from Harley Balzer as chairman of the executive committee of the International Science Foundation (ISF), the body set up by Hungarian financier George Soros to provide financial assistance to scientists in the former Soviet Union (FSU).

One of Watson's first tasks will be to secure new funding for the foundation, following Soros's decision that he will only provide further funds (in addition to the \$100 million he has already donated) where these match money raised from elsewhere.

In the short term, the ISF is still hoping to receive a grant from the US State Department sufficient to keep its grants programme running when Soros's money runs out. The longer term remains uncertain, since the US government has indicated its reluctance to spend large amounts of money in support of scientists in other countries.

As a result, Watson is said to be already engaged in a search for alternative donors, although one option is to try to persuade influential Congressmen to put pressure on the State Department to increase its commitment.

One issue concerning the foundation is the shortage of current scientific journals in FSU states, a situation which makes it virtually impossible for scientists to carry out effective research. The ISF has agreed to provide \$4 million to address this problem, and a first shipment of journals is due to arrive in Russia in the near future, to be distributed to the leading libraries.

Another scheme supported by the foundation is a \$5.5 million project to keep scientists in FSU states in touch with those elsewhere through satellite communications. Telecommunications systems connecting Moscow to Stockholm and Washington are currently being installed to reduce the cost of electronic mail for Russian scientists.

ISF's travel grant programme has already proved successful. The foundation has sent about 1,500 FSU scientists to foreign conferences, at a cost of \$2 million. Its executive committee has now decided to extend this programme and has allocated a further \$5 million to it.

Meanwhile several other programmes aimed at supporting science in Russia — some drawing directly on the ISF's expertise — are soon to start operation. These include the Howard Hughes Medical Institute, which plans to award 100 research grants worth \$100,000 each. Fifty will be awarded for biomedical research carried out entirely in FSU states, and the remaining 50 for collaborative research projects with US scientists.

Germany's Volkswagen Foundation has also announced plans to support joint Russian-German research projects, with individual grants also of \$100,000. In a separate

initiative, the Dutch government has launched a scheme for financing research involving Dutch and Russian scientists. The British government has already had such a scheme in place for some time.

Apart from the ISF, however, the largest source of financial support is the fund set up by the European Community (EC). Originally known as the Mitterrand Foundation, as it was created at the suggestion of the

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

French president, this has now been renamed INTAL (an abbreviation for the International Association for the Promotion of Cooperation with Scientists from the Newly Independent States of the FSU).

James Watson: seeking new funding.

The association has been given 21 million ECUs (US\$23 million) up to the end of this year to support joint projects involving scientists from FSU states and those from at least two European countries (purely bilateral collaboration is being handled at a national level).

China counts the cost of higher education

Beijing. As a new academic year opens, complaints are running high in China over the costs of university education. In the past, the government paid for everything, from tuition to accommodation, and guaranteed every student a job on graduation. But neither practice remains in force.

The State Education Commission has introduced a series of reforms to the country's higher educational system. Each university can now charge students for tuition, a move intended to ease the financial burden of the central government.

But as a result of these reforms, an increasing number of students are having to finance their own university education. According to the Shanghai Educational Test and Enrolment Center, over 6,100 self-supported students will be enrolled in Shanghai alone this year, just under 40 per cent of the total study enrolment.

Tuition fees vary. In Shanghai, self-financed students in liberal arts are being required to pay 2,500 yuan for one academic year, engineering students 2,700 yuan and arts students 3,000 yuan. The figures in Beijing are much larger, with an average annual tuition cost of about 4,000 yuan.

Such costs are high for many ordinary Chinese families. In a country where higher education can be the only path to a successful career, every means is being explored to

A deadline of 15 October had originally been set for the receipt of declarations of intent for all this money. But after concern had been expressed about the time needed for information about the programme to be disseminated — particularly to researchers working far from Moscow — the schedule has been revised. As a result, applications for half of the money still have to be submitted by the original deadline. But the executive committee of the association has agreed to set a second deadline at the end of January for the rest.

Meeting at the end of September, the association also agreed to accept two non-EC countries, namely Austria and Switzerland, as members. No budget has yet been set for next year, although officials say that the European Parliament will agree to a figure that is at least as high as this year's.

Vladimir Pokrovsky

■ In a recent article on President Yeltsin's decision to increase the stipends paid to members of the Russian Academy of Sciences (see *Nature* 365, 379; 1993), it was stated that the value of the new stipend of 150,000 roubles a month was worth the equivalent of US\$1,100. This figure should have been \$110.

secure the money needed. But many students from rural or frontier regions cannot afford to pay the high fees for their university education, and as a result the door of higher education seems closed to them.

This situation has prompted the State Education Commission to take action. In a circular issued jointly with the Ministry of Finance, it has asked universities to help students in particular need, for example by reducing tuition fees and granting of loans.

Some universities have already come up with their own strategies to cope with the challenge. For example, Southeast University located in Nanjing, the capital city of South China's Jiangsu Province, has introduced three ways of treating newly enrolled students, based on academic performance.

Those students in Category A, judged by their scores on a national college entrance examination, will receive full scholarships covering tuition and living. Students in category B will have tuition fees paid, but have to pay living expenses. Category C students cover their education entirely on their own.

These three categories will be rearranged each term according to the students' performance. But critics still argue that additional action needs to be taken to ensure that students from poorer families are not disadvantaged by being judged purely on their academic achievements.

You Qin Li

Japan digs deep for secrets of the Sun

Kamioka, Japan. A vast chasm 40 metres in diameter and 40 metres deep is being cut out deep under a mountain in the centre of Japan to accommodate a detector which, its designers hope, will answer one of the key questions facing modern cosmologists: why does the number of neutrinos observed from the Sun seem to be about half that predicted by theory.

The world's largest neutrino detector, the Super-Kamiokande observatory, is currently under construction 1 km down a zinc-and-lead mine known as the Kamioka mine. The detector is being built with total funding of grant of ¥10 billion (\$95 million) from Japan's Ministry of Education, Science and Culture, and should be ready for testing in two years time.

When neutrinos enter a tank of water, they occasionally collide with water

molecules, creating very faint Cerenkov radiation which can be picked up by photomultiplier tubes in the sides of the tank. Using such techniques, the smaller Kamiokande observatory in the same mine made headlines in 1987 when it detected neutrinos from a supernova in the Large Magellanic Cloud.

The new detector, Super-Kamiokande, will be more than 10 times the size of Kamiokande. Containing 50,000 tonnes of water, it should be able to detect about 30 solar neutrinos a day instead of only one every three days caught by the existing detector. Yoji Totsuka of Tokyo University's Institute for Cosmic Ray Research, who is leading the project, says that this larger detection rate will enable the observatory to help resolve the 20-year controversy over solar neutrinos.

The detector is located deep in the mine to filter out cosmic radiation that would disturb the observation of neutrinos. It runs largely automatically, recording events on magnetic tape.

But operating it still requires dedication from scientists. Kamiokande is reached by a 15-minute ride in an old tram down a slanting mine-shaft in pitch darkness. One scientist from Totsuka's group, assigned for an eight-day shift, monitors the machine during the day with only one tram in the morning and one out at night.



Under construction — the Super-Kamiokande observatory.

Just over a quarter of the construction money will be spent on the manufacture of 11,200 highly sensitive 20-inch photomultiplier tubes that will line the walls of the giant detector. Totsuka says he was able to negotiate a 60 per cent discount on the normal price of the tubes from Hamamatsu Photonics, a small Japanese company that is the world's only manufacturer of such tubes.

A similar sum is being paid to Kamioka Mining and Smelting Company, owner of the mine, for excavation and construction work. The contract is partly a reward to the company for supporting the earlier Kamiokande project which was built on a shoestring budget of only ¥500 million.

Excavation of the cavern to accommodate the detector will be completed by next August. The detector itself, as well as a water purification system and other facilities such as a network of workstations for collecting and analysing data, should be in place by October 1995, in time for routine observations to begin in the spring of 1996.

David Swinbanks

UK plans library of DNA profiles

London. Britain's Home Secretary, Michael Howard, announced last week that the government intends to establish a library containing the DNA profiles of virtually anyone convicted of committing a criminal offence.

The move, which will put DNA profiling on the same level as conventional fingerprinting, goes considerably further than the recent recommendation of the Royal Commission on Criminal Justice. This had suggested that a national DNA library should be restricted only to those convicted of relatively serious criminal offences, such as burglary and assault.

The Royal Commission's suggestion would have led to a library of about 180,000 profiles. The government's decision to go for a larger grouping will result in one containing 500,000, almost one in ten of Britain's population.

The government's decision was one of 27 new measures announced by Howard last week at the Conservative Party's annual conference in Blackpool designed to demonstrate the party's commitment to law and order. "DNA will be for today's policeman what fingerprints were for Dixon of Dock Green", said Howard, referring to the hero of a popular television series in the 1950s and 1960s.

Critics of DNA profiling, however, continue to warn of the need for caution in introducing its widespread use while questions about the accuracy and statistical validity of the tests remain in dispute. □

Mine finds uses for "inner space"

Kamiokande and Super-Kamiokande are not the only scientific experiments being carried out in the Kamioka mine, one of the largest and most modern zinc-lead mines in the world. The rapid rise in the value of the yen has reduced annual sales from about ¥30 to ¥17 billion. As a result, the company that operates the mine has been looking at a number of ways of exploiting it for scientific purposes.

In one experiment, an explosives company is renting a tank of water for carrying out underwater tests on new explosives. And a scientist from Tokyo University is helping to develop a large chamber for carrying out controlled tests on athletes under high-altitude conditions.

The athletes are sealed in the chamber, and sufficient air is sucked out

to simulate high-altitude conditions up to the equivalent of 4,000 metres. The athletes use exercisers while their physiological condition is monitored.

The high-altitude chamber, built at a cost of ¥40 million, is still at the test stage. But officials at the mine are hoping that companies interested in supporting Olympic athletes will invest money in a larger facility.

At present, such scientific activities — including the construction of Super-Kamiokande — account for about five per cent of the company's income. But with the mine only expected to be viable for about another 15 years, officials hope that what they describe as "inner space technology" will provide a growing source of revenue.

D.S.

Nobel goes to discoverers of 'split genes'

London. 'Split genes' have taken the 1993 Nobel prize for physiology or medicine. The prize has been awarded jointly to Richard J. Roberts, who now works at New England Biolabs, and Philip A. Sharp of the Massachusetts Institute of Technology.

At a meeting held at Cold Spring Harbor in June 1977, both reported independently that a single adenovirus messenger RNA molecule corresponded to four distinct regions of the DNA encoding it.

Up to that meeting, the general consensus was that genes were continuous stretches of DNA which served as direct templates for messenger RNA molecules, and that these molecules were themselves templates for protein synthesis. Belief in this model was understandable, as all studies of prokaryotic organisms had supported that theory.

But Sharp and Roberts showed that, at least in the adenovirus, things were not always so straightforward. Following the disclosure of their results, many other scientists soon reported that the interruption of coding sequences was found in other genes, and indeed that a discontinuous gene structure is the most common structure found in eukaryotes.

Pierre Chambon, of the Institute of Biological Chemistry in Strasbourg, was one of those who reported similar organizations of another gene, namely that for chicken ovalbumin. "There was no question that Phil Sharp would win the Nobel prize some day," says Chambon.

Chambon claims that Sharp "deserved the award on more than one account, not only for the 1977 discovery, but also for his later work". But he also stressed that, in the case of both Roberts and Sharp, the work was that of a team, and expressed regret that a Nobel prize cannot recognize this fact.

David Baltimore, at the Rockefeller University in New York, whose Nobel prize-winning discovery of reverse transcriptase also broke with the "central dogma" on the relationship between DNA, RNA and proteins, said that the award of the prize was long overdue.

"Everything that's happened in molecular biology since has depended on their discovery of split processing of mRNA," says Baltimore. "It has become a central focus of everybody's thinking on the field."

Sharp in particular has been a "notable contributor" for his broad contributions to the characterization of splicing intermediates of mRNAs, of transcription control, and of both basal and specific transcription machinery.

There was more praise from Walter Gilbert of Harvard University, who shared the 1980 Nobel Prize in Chemistry with Frederick Sanger and Paul Berg for his work on DNA sequencing. Gilbert says he re-

members vividly the dramatic reports delivered at the Cold Spring Harbor meeting. "The discovery of RNA splicing was radically new," he says. "There was a paradigm shift when Roberts and Sharp showed that the structure of genes was not what was expected."

Gilbert says that the basic discovery would have emerged for all to see with the advance of various techniques, such as nucleotide sequencing. But Roberts and Sharp had realised what was happening before anyone else. Other groups, he points out, had been looking at gene structures, but had failed to understand them; Roberts and Sharp made their discovery by 'seeing' the splicing of RNA through electron microscopy.

Following the discovery, Gilbert himself began to pursue the evolutionary implications of the finding that genes are discontinuous. It is now widely believed that this exon/intron structure of genes may drive evolution by allowing new proteins to be made by reshuffling different coding regions of the DNA. There are many examples where alternate splicing of one precursor RNA can lead to the production

of many different proteins.

Sir David Weatherall, regius professor of medicine at the University of Oxford, whose work has focused on the molecular pathology of human diseases — particularly of thalassaemias, some of which are due to defects in splicing of the globin genes — puts the work in its medical perspective.

"The basic biology is fascinating," says Weatherall. "But in terms of the molecular level underlying disease it has opened up a whole new area, not just in understanding the basic mechanism but also the variations of genetic disease."

Sharp himself describes the award as recognition of a "real discontinuity in science, to which many people contributed". He points out that soon after his and Roberts's announcement there were a "whole slew" of other papers that described spliced genes. "We were just lucky to be able to be at the beginning of the process", he says.

Like Chambon, Sharp also stressed that both his and Roberts's work was that of teams, and that all the members of their respective teams made significant contributions to the discovery.

Kimberly Carr

France urged to change ethics rule

Paris. The French national bioethics advisory committee this week recommended reform of laws on medical confidentiality, data protection, and free and informed consent, to bring regulations in these areas closer in line with the needs of psychology research.

The committee was convened earlier this year after the government had suspended a study on the cognitive traits of half-brothers and half-sisters born by artificial insemination from anonymous donors (see *Nature* 361, 481; 1993). This followed allegations in the French press that the researchers had used confidential medical data in the study, and had not provided sufficient details of the experiments when asking the parents of the children for their consent.

Biological experiments on humans and clinical trials are tightly regulated in France under a 1988 law. But the ethics committee says that the legislature has failed to consider the particular issues raised by behavioural studies on humans. For example, it says the legal requirement that subjects give free and informed consent is inappropriate to some psychology experiments, as prior knowledge of the objectives can modify a subject's response.

The committee reaffirms that subjects must give their free and informed consent to all experiments. But it recommends that the law be changed to let behavioural researchers hide some experimental objectives from

a subject in the interests of the experiment when requesting consent, on condition that they inform subjects of this.

Subjects would also be told that they could withdraw from the study any time, and that they would be given a complete "debriefing" on the full objectives and results of the study on its completion. If the subjects could be identified from the data collected, the researcher would also have to ask their consent for subsequent use of such data. Furthermore, the committee recommends that the need to hide information from subjects be approved beforehand by new bodies called "Consultative Committees for the Protection of Persons in Behavioural Research". These would be modelled on the committees that now approve clinical trial protocols.

The committees would also evaluate the "scientific pertinence of the research", and check that "the liberty, dignity and safety of subjects" is protected. Researchers would inform subjects before experiments that the committee had approved all hidden objectives.

The ethics committee also recommends that physicians should be allowed to share medical secrets with psychology researchers. At present it is illegal for physicians to tell anyone except another doctor about a patient, and then only if the information is used for therapeutic purposes.

Declan Butler

CNRS launches new strategy to break with its past

Paris. Europe's largest fundamental research organization, France's Centre National de la Recherche Scientifique (CNRS), last week launched an ambitious plan to increase its operating flexibility, continue its decentralization away from Paris, and build closer links with universities and industry.

Each of these represents a major step towards the "liberalization" of the organization and a shift away from past operating practices. But how effective the plan will be depends on the ability of both the CNRS and the government to overcome obstacles to change, in particular the legal conditions under which the CNRS operates and the terms of the employment of its scientists.

The plan, which outlines the CNRS's strategy for the period up to the end of 1995, was published in Paris last week. One commitment is to increase support for several interdisciplinary themes — such as understanding complexity — and raise its efforts to integrate mathematics, computing, and new technologies into all scientific disciplines (see box).

The goals of the plan are mainly drawn from last year's Rapport de Conjoncture, a 500-page review of international trends in research based on a wide-ranging consultation with CNRS scientists. A new element in CNRS thinking is the official encouragement being given to scientists to take more interest in the social and economic implications of their research.

François Kourilsky, director-general of CNRS, says it has become essential for researchers to consider the social accept-

ability of technology if they are to avoid a public backlash against their work. "We are certainly in favour of the fundamental research needed to develop new technologies, but not at any price", he says. "Technology must be adapted to man's needs, not the reverse."

Kourilsky says he would like to bring researchers together into interdisciplinary programmes with an explicit social dimension. In particular, he wants to see more social scientists, who make up one-sixth of CNRS research staff, participating in experimental research programmes. "I wish some of them would spend less time studying the sociology of the thirteenth century and more time on today's problems."

The emphasis being placed by CNRS on the relevance of its work to social issues is partly intended to reassure politicians and public of the value of basic research, particularly at a time when France's economic crisis is biting into public spending. "We have to defend research", says Kourilsky.

For the same reason, CNRS's new strategy emphasizes the importance of basic research in creating technology and jobs, pointing out that it has increased the value of its contracts with industry from FF39 million (US\$7 million) in 1983 to FF550 million last year.

CNRS has also, for the first time, been consulting with industry over its choice of research priorities. For example, it accepted the advice of industries based on the life sciences not to focus excessively on human genetics at the expense of more traditional fields such as basic microbiology.

Perhaps the most radical aspect of the new strategy is the strong encouragement to be given to research workers to take up either temporary or permanent positions in universities. CNRS may have little choice. François Fillon, the minister for higher education and research, is expected to stipulate targets for such transfers within the pro-

French labs face funding squeeze

Paris. The small increase in the 1994 French budget for science (*Nature* 356, 378; 1993) will help the immediate cashflow of the Centre National de la Recherche Scientifique (CNRS). But it will seriously restrict any new initiatives next year.

CNRS says that the budget translates into a 2.5 per cent cut in operating expenses, a 3.7 per cent cut in programme authorizations and a 3.8 per cent increase in actual cash payments. These compare to the anticipated rate of inflation of 2.2 per cent.

The increase in cash payments will let the

CNRS make up most of last year's shortfall. But the cut in programme authorizations will mostly fall on laboratories, says CNRS, because it cannot reduce its commitments to international big science projects.

Plans for a large increase in foreign postdoctoral fellowships have also been frozen. Indeed, some French researchers are concerned that the CNRS will not be able to continue to provide additional support to laboratories that have already agreed to take on foreign postdoctoral fellows.

Declan Butler

The CNRS strategic plan, 1993 – 1995

Science:

- Understanding complexity (from molecules to mathematics)
- Development of technologies and processes better adapted to human needs
- The main challenges facing society (the environment, killer diseases, ageing, cities, unemployment)
- Big science (astronomy, particle physics)
- New tools and new technologies for research (large equipment, mathematical modelling, electronic highways, databases, nano-technology, robotics)

Policy:

- Greater collaboration with other research organizations and industry
- Decentralization and internationalization
- Dissemination of science in society
- More active personnel management
- Merger of groups of laboratories into semi-autonomous research institutes

posed five-year contracts between his ministry and the CNRS and other research organizations (see *Nature* 365, 196; 1993).

Only 40 or so CNRS researchers transferred to universities last year. Kourilsky favours greater mobility in both directions, and has already increased incentives for the agency's researchers to take on teaching duties.

He also sees Fillon's demand for greater collaboration with universities as a solution to the staff logjam within the CNRS. The agency spends three-quarters of its budget on salaries, leaving little money for research funding. The small growth in next year's budget has narrowed the margin further.

Kourilsky admits there is a strong case for reducing staff numbers. But he points out the difficulty of doing this, given that all CNRS employees are civil servants, and therefore enjoy full job security. His plan is to reward the best researchers with rapid promotion, and provide incentives for the less productive ones to take up teaching posts in universities or to enter other public services.

Roland Tournereau, from the researcher's union SGEN-CFDT, says that the union generally supports such moves. But CNRS has little experience of the sort of case-by-case personnel management that implementing this policy would require. And given overall budgetary constraints, its success in doing so — as in shifting resources into priority areas — also depends on whether Fillon is able to secure changes in legislation to give CNRS greater responsibility to manage and spend its budget.

Declan Butler

US-Ukraine stalemate over nuclear weapons

Washington. Ukraine's opposition to the complete relinquishment of strategic nuclear weapons located on its soil is hardening, and the process began even before the latest Russian crisis increased its security concerns, according to US representatives recently returned from two meetings on disarmament held in Kiev.

Indeed some observers fear that international concern about the fate of the 170 intercontinental missiles which remain under Russian control on Ukrainian territory may be counter-productive. Officials and politicians in the troubled republic are being drawn to the issue, says one disarmament expert, because the more belligerent their stance, the more attention they get.

The meetings were organized in the Ukrainian capital late last month. One, sponsored by the American Association for the Advancement of Science (AAAS), focused on the disposal of nuclear weapons in Ukraine itself, while a United Nations symposium addressed the wider question of disarmament across the former Soviet Union.

One concern raised at both meetings was that the United States — in common with France and the United Kingdom — may not be in a good position to advise Ukrainians that nuclear weapons are a costly and ineffective way of meeting their defence needs.

Another concern is whether the West is either able or willing to meet Ukraine's demands. The United States wants it to sign the international non-proliferation treaty as a non-nuclear state — the nuclear weapons on its territory are not under its control — and to comply with US-Soviet agreements on strategic weapons reductions by sending the weapons to Russia for dismantling.

But Ukraine says that, before doing so, it needs security guarantees from the West, money for the fissile materials it sends to Russia, and more money to help with the dismantlement process and the redeployment of up to 80,000 soldiers associated with the weapons sites. The United States is said to be offering \$175 million for the

entire process; Ukraine wants \$3 billion.

Ukraine says that it has no intention of seeking operational control over the missiles on its territory, and the US officials say that this desire appears to be genuine. Marshall Brown of the US Arms Control and Disarmament Agency, who attended the AAAS meeting, says that in his personal view "realists" in Ukraine are "losing interest in the retention of nuclear weapons".

But James Doyle, an analyst with Science Applications International Corporation who also attended the AAAS forum, believes that "the trend is in the opposite direction". Doyle fears that a dangerous stalemate will continue.

The United Nations forum in Kiev does not appear to have been any more hopeful. "The mood is moving in favour of retention, not relinquishment", says one US attendee. With turmoil in Russia, and the West's interest in Ukraine apparently focused on this single, thorny issue, the prospects of a speedy resolution are not bright.

Colin Macilwain

Ig Nobel prizes reward fruits of unique labour

Boston. On the night of 7 October, a handful of Nobel laureates and hundreds of science 'enthusiasts' descended on the Massachusetts Institutes of Technology (MIT) for the announcement of the 1993 Ig Nobel prizes. These prizes, which are as feared as their Stockholm counterparts are coveted, are awarded for research that "cannot or should not be reproduced".

The event, which was being held for the third time, is sponsored by *The Journal of Irreproducible Results* (JIR) and the MIT Museum, and the evening's festivities were hosted by the journal's editor Marc Abrahams, Sheldon Glashow (Nobel laureate in physics, 1979) and William Lipscomb (Nobel laureate in chemistry, 1976). 'Ig Nobels' were presented in 11 categories, although not all the "winners" saw fit to attend the ceremony. "I don't see how this could possibly help me or my company", one of the absentee ignobelists explained.

The prize for psychology went to John Mack of the Harvard Medical School and David Jacobs of Temple University for lending credence to the notion that thousands, if not millions, of people have been abducted by aliens from outer space. Neither Mack nor Jacobs was there to accept the honour, but an alleged assistant attorney general took the stage, reminding the audience that "kidnapping is still a federal offence".

Paul Williams of the Oregon State Health Division and Kenneth Newall of the Liver-

pool School of Tropical Medicine captured the Biology award for their pioneering study, "Salmonella excretion in joy-riding pigs". The prize for literature went to the 972 co-authors of a *New England Journal of Medicine* paper published in this year's 2 September issue. Journal editor Marcia Angell stood in for the 972 heart specialists, describing the division of labour: each co-author was responsible for exactly two words.

Robert Faid of Greenville, South Carolina made even more elaborate calculations, showing that the odds of Mikhail Gorbachev being the "Antichrist" are approximately 8,606,091,751,882 to one. Faid's efforts were rewarded, not surprisingly, with the mathematics 'Ig'. French researcher, Louis Kevran, won the physics prize for his conclusion that the calcium in a chicken's eggshell is created by cold fusion.

The final 'Ig', in the field of medicine, went to James Nolan of the Guthrie Clinic in Sayre, Pennsylvania, and two colleagues, for the painstaking report "Acute management of the zipper-entrapped penis", published in *The Journal of Emergency Medicine*.



Demonstrators protest against the planned destruction of subatomic particles in the Superconducting Super Collider

The 1993 ceremony included a new feature, "The Heisenberg Certainty Lectures", presented by a number of luminaries. Alan Lightman (author of *Einstein's Dreams*) delivered his lecture in Pig Latin. Sheldon Glashow spoke in English, although he was frequently interrupted by members of the PLO (Proton Liberation Organization), protesting against the Superconducting Super Collider. "Billions of subatomic particles will be threatened every picosecond," one protester shouted. "Stop the carnage!"

On this night, however, another heckler had the final word. At the end of the ceremony, a man yelled from the balcony, "What a waste of time!". Perhaps, but many felt it was time well wasted. **Steve Nadis**

Setting the record straight

SIR — The review of *Equivalence and Priority: Newton versus Leibniz* by William R. Shea (*Nature* **364**, 681; 1993) contains serious factual misrepresentations of my work on Leibniz and Newton's *Principia*. Shea claims that the new evidence in my book rests on a reading of just two manuscript sheets by Leibniz, together with the fact that one of them has a watermark of a type used by Leibniz in Vienna in 1688. On the basis of this "slender evidence", he says, I conclude that Leibniz saw the *Principia* in that city in that year and overlook the fact that Leibniz could have easily carried old paper with him and used in 1690 a sheet acquired a couple of years earlier. Hence my belief, contrary to Leibniz's own claims, that he saw the 1687 *Principia* before publishing his reply in 1689 would be unjustified.

This is not so much a charge of poor scholarship as one of lack of common sense. However, on p. 96 of my book I outline the criteria I have adopted for dating Leibniz's manuscript essays. After stating that "all watermarked manuscripts discussed in this chapter are of a type used by Leibniz in Vienna in 1688", I add that "this element alone does not prove that the essays date from 1688, since Leibniz could have carried some Viennese paper with him". I also add that my "second criterion involves the conceptual development of Leibniz's theory and terminology employed, and provides a firmer basis for the dating". Indeed, my work is based on an extensive investigation of several dozen manuscripts by Leibniz which are listed in Appendix 3. Moreover, I publish six of them for the first time, discussing the issues of priority and publication in a seventeenth-century context, and providing an extensive analysis of the background, formation, and reception of Leibniz's response to Newton.

Domenico Bertoloni Mell

Jesus College, Cambridge CB5 8BL, UK

Allain defended

SIR — Badrig Mélékian (*Nature* **365**, 289; 1993) repeats two of the myths of the "Paris blood scandal". He asserts that Dr Jean-Pierre Allain deliberately provided patients with infected material. This is untrue. Allain's problem at the beginning of 1985 was in convincing others of the urgency of the situation; haemophiliacs were found to be seropositive against HIV, but the full significance of this seropositivity in terms of infection was not clear at that time either to him (see his paper in *Blood* **66**, 896–901; 1985), or to

others (see the leading article by Montagnier *et al.* *Presse Médicale* **14**, 1451; 1985). The active steps Allain took to speed the introduction of safety measures in the period January–June 1985 have been detailed in the *Lancet* (**342**, 332–333; 24 July 1993). More recently, letters have arrived from Professor Gluckman and Dr Barré-Sinoussi of Paris who were as well-informed as anyone in the world at that time; both tell of the careful and concerned consultations that took place in early 1985 that led to the effective measures of June 1985.

Between June and October 1985, non-heat-treated pooled plasma continued to be given to seropositive haemophiliacs. But this was the result of an independent decision of the National Advisory Committee on Blood Transfusion (CCTS) made on the basis of the best knowledge of that time. Even in retrospect, this decision seems not unreasonable; furthermore, it was a consensus decision of first-class professionals and did not involve Allain.

There were not "hundreds and maybe thousands" infected during this period; the great majority of infections occurred before the end of 1984, before there was an awareness of the problem and before protective measures were available. That any infections at all occurred is a matter of profound regret, none more so than to Jean-Pierre Allain who is known to all who have worked with him as a caring and thoughtful physician, both before, during and after the events of 1985.

Robin Carrell

*University of Cambridge,
Department of Haematology,
MRC Centre, Hills Road,
Cambridge CB2 2QH, UK*

Korean women

SIR — In response to your article on the rapid development of science and technology in South Korea (*Nature* **364**, 379; 1993), I should like to mention the position of Korean women in those fields.

I grew up within the major science complexes in Korea. In 1980, when my father was president of the Korea Advanced Institute of Science and Technology, an application for a faculty position arrived from Massachusetts Institute of Technology. The application was taken seriously until it was realized that the applicant was a woman. In his whole career, my father has had only one female PhD student, and he had difficulty in finding her a job because there was a regulation against hiring female scientists in industry. The ban has since been lifted, but little else has changed.

Discrimination begins at a very early stage, as female students are often discouraged from pursuing careers in science (or any career at all, for that matter). If

the situation is changing, the changes are occurring very slowly. It is a pity that the country disregards the potential and talents of half the population when it is so concerned about improving science and technology. I have watched the talents of many bright and capable women go to waste — my mother's friends included unemployed PhDs whose frustration was apparent.

Carol Lee

*School of Oceanography WB-10,
University of Washington,
Seattle, Washington 98195, USA*

When in Rome

SIR — I should like to comment on "Know your Japanese host" (*Nature* **364**, 376; 1993). There are at least two necessary conditions in order to profit from a long stay in a foreign country: (1) understanding of the language of the host country, and (2) interest in its culture.

The primacy of the English language in the scientific community is undeniable, but English-speaking scientists will be at a disadvantage if they set out without consideration of the above factors, not only in their professional activity but also in their daily life in a foreign country.

It is not always easy to master a foreign language, but Japanese is not quite such a difficult language as your article suggests. The difficulty of Japanese for the English should be exactly the same that of English for the Japanese. Most Japanese scientists work hard to learn foreign languages, not only English, before they go abroad. It is neither fair nor useful to blame the lack of ability in English of Japanese students (*Nature* **362**, 387; 1993) without mentioning the inability of students from elsewhere to speak Japanese.

It is certainly true that "international exchange between Japanese and foreign researchers would benefit both groups" (*Nature* **340**, 337; 1989) but this may be achieved only by effort on both sides, based on fellowship and mutual understanding of cultures.

Nobuo Terao

*Catholic University of Louvain,
21 rue de l'Angélique,
B-1348 Louvain-la-Neuve, Belgium*

Prudent fellow

SIR — It's difficult to say this diplomatically, but, judging by the tone of Stuart Sutherland's review of *B. F. Skinner: A Life* (*Nature* **364**, 767; 1993), perhaps Skinner's avoidance of the edge of Magdalen Tower was, in the circumstances, not so much acrophobia as good sense.

Robert E. Adler

*39560 Stevenson Place, Suite 215,
Fremont, California 94539, USA*

Unpackaging and repackaging photons

The well-known inconvenience that a measurement on a quantum system will not leave it unchanged can be circumvented by turning a photon stream into an electrical current and then back again.

THESE days we all know that a measurement made on a quantum system inevitably interferes with the state of that system. Familiar examples, many going back to the 1920s, illustrate the phenomenon. There is, for example, the case of Bohr's microscope for measuring the position of a scattering particle in its field, originally intended as a way of demonstrating the Uncertainty Principle.

More recently, the general principle that a measurement will disturb the quantum state of the system measured has become the foundation of the doctrine of quantum measurement — not to mention the inspiration of the still more recent attempts to make measurements inherently more precise than those prescribed by Heisenberg. The catch-word for the second endeavour is QND, for "quantum non-demolition" measurement, where the implication is that the original quantum state is not, or not entirely, demolished by the process of making a measurement.

At the outset, the chief beneficiaries of QND would have been the small band of those who seek to use mechanical detectors to pick up the signals of gravitational waves spreading through the Universe from distant gravitational cataclysms. How, for example, would they distinguish vibrations of the detectors set in train by gravitational impulses from those occasioned simply by local thermal noise?

Now that light-beams routinely carry masses of information through the world's networks of optical cables, there are more immediate benefits to be won from QND, notably that it may be possible to extract information represented by the modulation of such a beam without destroying the original, perhaps so that the same information can be read out a second time, perhaps so as to design better repeaters (or signal-boosters) for optical cables.

Simple-mindedly, the task is straightforward. Why not use a photo-cell of some kind to turn the light-beam carrying information into an electrical current, amplify that with an amplifier of suitable frequency range (as free from inherent noise as possible) and use the resulting current, perhaps as the pump-input of a laser, to generate a light-beam that mimics the original?

Only a little reflection will show that this scheme would not always work well. For one thing, the signal-to-noise ratio in the original beam will be degraded by the efficiency of the two conversion processes and, more seriously, there will be an extra source of noise in the output beam because, how-

ever efficient the photomultiplier, the stream of electrons it generates will consist of discrete quanta of electrical charge whose distribution in time is to some extent statistical. "Shot-noise" is the vivid term for the phenomenon, loss of phase information is its consequence.

There is a further complication because light-beams themselves consist of streams of photons whose distribution is again statistical, at least in the regime where quantum effects matter (where the intensity is sufficiently great). That is again a form of shot-noise. Those concerned with the extraction of meaningful information from light-beams are as likely as not to be bothered by that difficulty.

Luckily, there may be ways round the difficulty. Two groups, one from Sweden and one from France, seem now independently to have shown that it is possible to extract information from a light-beam and yet regenerate a copy of the original beam carrying essentially the same information. Each group's experimental protocol rests on the use of semiconductor light-emitting diodes (LEDs) for converting an electrical current back into a light-beam. The groups are Edgard Goobar, Anders Karlsson and Gunnar Björk from the Royal Institute of Technology at Stockholm and J.-F. Roch, J.-Ph. Poizat and P. Grangier from the Institute for Theoretical and Applied Optics at Orsay, near Paris (*Phys. Rev. Lett.* **71**, 2002 & 2006; 27 September 1993).

The technology of these experiments is essentially the same, even if the details differ. The input beam generates an electrical current in a photo-detector (cooled to liquid nitrogen temperatures for efficiency) which is then used for two purposes: to provide an estimate of the signal carried by the original light-beam and to regenerate a simulation of the original. In each arrangement of the equipment, there is some stage at which the electrical current generated by the photo-detector is amplified by a low-noise amplifier. The core of each device is a semiconductor LED (also cooled) for converting the electrical signal back into optical form.

Do these arrangements indeed function as QND devices? None of them describes the recovery of a signal representing the voice of the latest pop-singer and a proof that the same signal remains intact in the output from the LED, but that is forgivable in the circumstances. On the contrary, the proof that some measure of QND has been achieved is complicated, and boils down to

demonstrating correlations between the input and the measured (or "meter") signals and between the meter and the output signals. Both groups claim similar success on both counts. That the meter signal is at least an approximation of the information carried by the input beam is taken for granted, but evidently matters crucially to those concerned with the operation of telecommunications networks.

What the Swedish and French groups have done is in one sense a celebration of the convenience and relatively high efficiency of the LED devices now available. (Even so, the efficiency claimed by both groups is only about 30 per cent, whence the need that there should be some kind of amplification in the circuit.) Their work appears to be the latest of a series of exercises stretching back over the past five years to demonstrate QND in the manipulation of light-beams near the shot-noise limit.

The literature of the past few years includes systems in which two or three high-intensity laser beams are coupled together through non linear optical elements, systems in which the optical output of a LED has been directly coupled to a laser through a nearly transparent end-facet and semiconductor devices that combine the properties of photodetection and light-emission. It is reasonable to suppose that some industrial companies may already be hard at work on the design of chips of silicon that will extract information from a light-beam in an optical fibre working at high intensity without distorting the signal delivered to people further down the line.

What that implies is that QND measurements may be within reach of everyday application. Does that mean that Heisenberg's Uncertainty Principle can be discarded? Sadly for those who regard the uncertainty of quantum phenomena as an uncomfortable assault on intuitive notions of causality (which it is not), that is not the case. In the optical field as in other attempts at QND, the essential trick is to arrange that the interference with the quantum system engendered by a measurement appears only in the conjugate variable. In the experiments now reported, the price of extracting almost faithful information about the amplitude of the original light-beam is offset by the complete loss of register between the phase of the meaningful signal and that of the carrier wave in the output signal. Mercifully, for those who are interested only in the signal, that is an irrelevance.

John Maddox

In search of the halo grail

Craig J. Hogan

WE know that invisible 'dark matter' dominates the mass of the Universe. Something of its quantity and distribution is revealed by the effects of its gravity on the motion of visible stars, gas and galaxies¹, but there has been no direct evidence to show what dark matter is made of. On pages 621 and 623 of this issue, two groups^{2,3} report several seemingly ordinary astronomical events — three stars in the Large Magellanic Cloud that gradually brightened, then dimmed again, over a period of about two months — which could herald a breakthrough in revealing the composition of cosmic dark matter. The observed variation in stellar brightness is consistent with the magnification of stellar images expected from the gravitational effects of small dark stars moving across the line of sight in the foreground: that is, within the dark halo of our Galaxy.

Studies of the orbits of stars and nearby satellite galaxies⁴ show that the visible disk of our Galaxy is embedded in an extended dark halo, perhaps ten times as massive as the visible stars, but emitting no detectable light. One reasonable hypothesis⁵ is that the halo is made up of 'Jupiters' or dark 'brown dwarf' stars, cold stable objects too small to heat up to nuclear burning temperature, also called 'degenerate dwarfs'. Or it could consist of compact remnants of burned-out massive stars, such as neutron stars or black holes. These objects have been collectively dubbed 'massive compact halo objects', or MACHOs, as opposed to elementary particles such as massive neutrinos, photinos or axions. The key difference is that individual MACHOs have detectable gravitational effects.

Gravitational lenses

The gravitational field of a point mass bends light rays from a background light source. In the present case of 'gravitational microlensing', the bending angle is so small that no image distortion is seen. But like other lenses, a gravitational lens changes the brightness of a star image. Since the configuration of an astronomical lens changes with time (because of the relative motion of the observer, the source, and the lens), so does the magnification. If the galactic halo is made of MACHOs, the brightness of a star outside our Galaxy will occasionally vary as a MACHO passes close to the line of sight^{6,7}. These events are rare; only about one star in two million should show a detectable effect (brightening by more than 30 per cent) at any given time. To detect the effect, millions of stars must be monitored for variations.

The new discoveries result from two similar carefully planned experimental efforts mounted to detect this microlensing effect on images of stars in the Large Magellanic Cloud, the nearest galaxy to our own. One group used a 32-megapixel array of CCDs (charge-coupled devices) on a dedicated telescope at Mount Stromlo observatory in Australia²; the other used scanned, digitized photographic plates from the European Southern Observatory in Chile³. Culling with sophisticated software the thousands of images of millions of stars taken over hundreds of nights of observing, they found many interesting variable stars, and three good candidates for gravitational microlensing in the dark halo of our Galaxy. The time variation of the brightness of the candidates is very close to that predicted using the simple model of the passage of a point lens in front of a star, and the variation is the same at different wavelengths, which is not generally expected from intrinsic stellar variation. The duration of the events (one to two months) gives a reasonable estimate of the mass of the lenses, at between about 0.03 and 0.3 times the mass of the Sun (or up to one solar mass in the case of ref. 2). But such estimates are uncertain because of the many unknown details of the relative velocity and position of the lens.

With so few possible detections, the case for microlensing is not yet conclusive. But with the quantity of data being generated, there should soon be a statistical sample in hand — say, more than ten events — which will allow further tests. For example, the theory predicts the distributions of magnifications (larger being rarer), and predicts that the lensing is independent of the type of star being lensed. With a large statistical sample, one can start to measure properties of the mass distribution of the lenses. Perhaps the most exciting result in these two papers is that the large-scale monitoring project is practicable, that backgrounds can be dealt with, and that we therefore have a powerful new technique to probe the composition of cosmic dark matter.

Given the amount of data so far analysed, the number of events seen is about that expected if the halo dark matter is composed entirely of MACHOs. Suppose that over the next year or two a dozen or more similar candidates appear, confirming the hypothesis that the halo is made of MACHOs. What will that tell us?

We already know that the halo mass cannot be made up mostly of hydrogen-burning stars (those larger than about 0.08 solar masses), because they burn too brightly and would have been seen

already. On the other hand, the estimated mass ranges for the present candidates are too low to be neutron stars or black holes, at least for known production mechanisms for these objects. The most plausible interpretation is that the bulk of the material is in low-mass degenerate dwarfs, supported against gravity not by thermonuclear heat but by the degeneracy pressure of cold electrons. If this guess is correct, subsequent events should generally produce mass estimates smaller than the first candidates.

Cosmic baryons

Such dwarfs have long been a favourite candidate for the galactic halo dark matter, because of a simple need to account for cosmic baryons (ordinary protons and neutrons). For the simple 'hot Big Bang' theory of nucleosynthesis to predict the correct relative abundances of the light elements, we need the mean cosmic density of baryons, in units of the critical density, to be about $\Omega_b = 0.015h^{-2}$, where $h \approx 0.5 - 0.8$ is Hubble's constant H_0 in units of 100 km s^{-1} per megaparsec. But altogether the bright shining stars of the Universe contribute only about $\Omega_{\text{stars}} \approx 0.003h^{-1}$, so most of the baryons must be dark. The dark haloes of galaxies, extrapolating using the mass-to-light ratio of our Galaxy, have according to various estimates a mean density of $0.017h^{-1}$ or even more, and therefore might be suitable repositories for the 'missing baryons' of the Big Bang, if they are locked up in dark and practically invisible degenerate dwarfs. The microlensing programmes may at last reveal the form and hiding place of most of the baryons in the Universe.

The confirmation of haloes made of compact objects will strongly affect theories of galaxy formation and evolution. Normal star formation as it occurs today in the disk of the Galaxy does not lock up so much material in relatively small bodies, so a halo dominated by dwarfs requires the initial mass function to be different (steeper) in the past⁸. Many theories of cosmogenesis do not predict MACHO haloes; cold dark matter models, for example, generally make galactic haloes predominantly out of primaeval cold dark material such as axions or photinos (the extra unseen baryons hiding in a hot intergalactic medium, rather than in collisionless galactic haloes). If the halo indeed turns out to be made largely of MACHOs, then ongoing direct laboratory searches for primaeval weakly interacting massive particles (WIMPs, naturally enough) in the galactic halo should not find them.

Resolving the nature of the galactic halo may still be only the first step in solving the riddle of the composition of most dark matter in the Universe. If standard Big Bang nucleosynthesis is right, then all of

the baryons together, including the MACHO haloes, still make up only about $\Omega_b = 0.015h^{-2}$. In spite of the persistent uncertainty in h which bedevils this cosmic bookkeeping, this is probably not enough to provide all of the cosmic dark matter, the bulk of which ($\Omega \geq 0.1$) is found in the dynamics of galaxy clusters and superclusters on larger scales. This reasoning argues for at least two kinds of dark matter, degenerate dwarfs for galaxy scales and something else on larger scales. A popular option for the second kind would be enough massive neutrinos for a flat Universe, giving $\Omega = 1$.

Quasar microlensing might give another probe of the cosmic density of MACHOs, although at present these studies involve more indirect statistical arguments than the galactic halo events. Microlenses should occasionally amplify the small continuum-emission regions of quasars, but not the line-emitting regions because of their large angular size. The fact that few, if any, quasars show an anomalously small line-to-continuum ratio argues against a large global density ($\Omega = 1$) of objects in the mass range 0.001 to 300 solar masses^{9,10}. On the other hand, a recent statistical analysis of quasar variability¹¹ argues that cosmic MACHO microlensing is ubiquitous, requiring a mean cosmic density of at least $\Omega \geq 0.1$ in objects between 0.001 and 0.1 solar masses. In this case, dark matter may all be baryonic MACHOs. This situation can be reconciled with standard cosmology and with most data on cosmic structure^{12,13}, although reconciling such a large baryon density with Big Bang nucleosynthesis might require MACHOs to have formed very early in the Universe¹⁴. □

Craig J. Hogan is in the Astronomy Department, University of Washington, Seattle, Washington 98195, USA.

1. Binney, J. & Tremaine, S. *Galactic Dynamics* (Princeton Series in Astrophysics, 1987).
2. Alcock, C. *et al.* *Nature* **365**, 621–623 (1993).
3. Aubourg, E. *et al.* *Nature* **365**, 623–625 (1993).
4. Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C. & Aaronson, M. *Astrophys. J.* **345**, 759–769 (1989).
5. Lynden-Bell, D. & Gilmore, G. (eds) *Baryonic Dark Matter* (Kluwer, Dordrecht, 1990).
6. Paczynski, B. *Astrophys. J.* **304**, 1–5 (1986).
7. Griest, K. *Astrophys. J.* **366**, 412–421 (1991).
8. Lake, G. in *Trends in Particle Astrophysics* (eds Cline, D. & Peccei, R.) 43–52 (World Scientific, 1992).
9. Canizares, C. R. *Astrophys. J.* **263**, 508–517 (1982).
10. Delcanton, J. J. *et al.* *Astrophys. J.* (in the press).
11. Hawkins, M. R. S. *Nature* (in the press).
12. Cen, R., Ostriker, J. P. & Peebles, P. J. E. *Astrophys. J.* **415**, 423–444 (1993).
13. Peebles, P. J. E. *Principles of Physical Cosmology* (Princeton Series in Physics, 1993).
14. Hogan, C. J. *Astrophys. J.* **415**, L63–L66 (1993).

Erratum

In Aaron Klug's News and Views article of last week (*Nature* **365**, 486–487; 1993) there are four instances where a Greek beta wrongly appears instead of the letter B in referring to the B-DNA form (col. 3, line 22; col. 4, line 7; and lines 1 and 8 in the caption to Fig. 2).

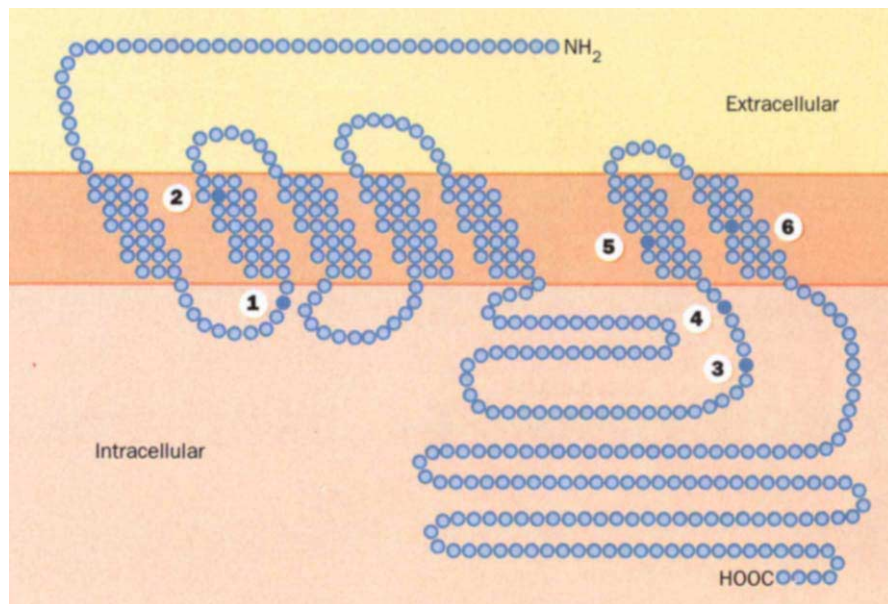
Turned on to ill effect

Robert J. Lefkowitz

ON pages 649 and 652 of this issue, Parma *et al.*¹ and Shenker *et al.*² describe the identification of the defects responsible for an uncommon form of hyperthyroidism and for familial precocious puberty. Most notably it turns out that both result from mutations in G-protein-coupled receptors which, in these cases, mean the receptors remain turned on. These are the

tors regulate levels of cyclic AMP (by stimulating adenylyl cyclase), either in the thyroid gland (thyrotropin) where the second messenger controls cellular differentiation and growth, or in the Leydig cells of the testes (luteinizing hormone) where it causes hyperplasia and production of testosterone.

Given the numerous physiological



Depiction of the proposed seven-transmembrane-spanning domains of a prototypic G-protein-coupled receptor, showing the location of spontaneously occurring mutations which cause activation. The mutations are found in the receptors for melanocyte-stimulating hormone¹² (1,2), thyrotropin¹ (3,4), luteinizing hormone² (5) and rhodopsin¹³ (6). The region mutated *in vitro* in the constitutively activated adrenoceptors^{9–11} overlaps that between points 3 and 4.

first such reports; others will probably follow.

Plasma membrane receptors allow eukaryotic cells to sense and respond to a wide variety of external stimuli, a feat accomplished by discriminatory ligand binding and receptor conformational changes which in turn influence the activity of downstream effector molecules. Normally, the transition to the active state is tightly controlled by interaction of the receptor with the appropriate stimulus. The two new papers provide examples of mutations in G-protein-coupled receptors which abrogate this control and make the receptors 'constitutively' active even in the absence of ligand. Parma *et al.* find that activated thyrotropin receptors, caused by somatic mutations of the receptor gene, produce functioning thyroid adenomas with hyperthyroidism. And Shenker *et al.* show that a rare autosomal dominant disease, in which young male children develop precocious puberty, is due to an activating mutation in the receptor for luteinizing hormone. These recep-

roles of G-protein-coupled receptors in neurotransmission, endocrine signalling, olfaction, vision and chemotaxis amongst others, it might be expected that abnormalities of such receptors underlie a variety of human pathologies. Indeed, several mutations that truncate or otherwise disturb the structure and functions of receptors for vasopressin³ and adrenocorticotrophic hormone⁴ have already been shown to be associated with states of hormone resistance. But the cases now reported have mutations that cause hyper rather than hypo function of the affected receptors. That is not a complete surprise, for the possible existence of such mutations was evident from *in vitro* mutagenesis studies^{5,6}.

The figure depicts a prototypic G-protein-coupled receptor. Regions in the intracellular loops have been shown to be involved in coupling the receptors to the G proteins. Interestingly, synthetic peptides — as short as 15 amino acids in length — that mimic those sequences can activate purified G proteins *in vitro*^{7,8}. So one

function of the intact receptor structure must be to prevent these small activating domains from interacting with the G protein until an agonist binds to the receptor. How this happens is not known, but presumably it involves intramolecular interactions which maintain the receptor in a constrained, inactive state until relieved by agonist binding.

Mutations in the carboxy-terminal segment of the third cytoplasmic loop of the $\alpha 1B$ adrenoceptor lead to constitutive activation of G-protein-mediated phospholipase C activation⁵. One residue was mutated to all possible residues and, remarkably, increased constitutive (that is agonist-independent) activity, compared to the wild type⁹, was observed in every case. Mutations of the cognate residues also activated the G_s -coupled $\beta 2$ adrenoceptor¹⁰ and G_i -coupled $\alpha 2$ adrenoceptor¹¹. The wild-type residues in these positions must therefore be uniquely suited to the task of keeping the receptor in its constrained or inactive conformation.

The mutant receptors described by Parma *et al.* and Shenker *et al.* display properties similar to those of the constitutively active mutant adrenoceptors created *in vitro*. In particular, when transfected into cultured cells both produce increased agonist-independent 'basal' levels of cAMP compared to wild-type receptors. The activating mutations in the thyrotropin receptor occur in the very same residues that activate the adrenoceptors *in vitro*, further underscoring the importance of this region in regulating the transition from inactive to active states of the receptor. For the luteinizing hormone receptor, the mutation is in an adjacent region (the flanking sixth transmembrane domain).

However, neither group examined the entire sequence of the receptor for mutations. So perhaps other activating mutations are present in distinct regions of the thyrotropin receptor in some of the other (8/11) patients with thyroid adenomas in

which no mutations were found in the third cytoplasmic loop (the region analysed by Parma *et al.*). It is worth noting that point mutations in the second transmembrane domain and first cytoplasmic loop of the receptor for murine melanocyte-stimulating hormone, which lead to darkened coat colour, cause constitutive activation or hyper-responsiveness of the receptors¹². In the case of rhodopsin, a point mutation of lysine at position 296 in the seventh transmembrane span, which prevents attachment of the chromophore retinal, leads to constitutive activation of rhodopsin signalling *in vitro* and is associated with retinal degeneration¹³. This suggests that retinal keeps rhodopsin in an inactive conformation, and in this sense acts as an endogenous negative antagonist — that is, as an agent that stabilizes the inactive conformation of a receptor. Thus there are likely to be a number of regions of G-protein-coupled receptors, as well as sites in the third cytoplasmic loop, which are essential in keeping the receptor in a constrained, inactive conformation. Such regions probably serve as key loci for

activating mutations (see figure).

Several G-protein-coupled receptors that activate phosphatidylinositol turnover can serve as conditional proto-oncogenes, transforming cells in which they are overexpressed, but only in the presence of agonist¹⁴. Moreover, after mutational activation, the $\alpha 1B$ receptor displays agonist-independent transforming properties (that is, it functions as a true oncogene¹⁵). The study by Parma *et al.* is the first to document this mechanism in humans and to extend it to adenylyl-cyclase-coupled receptors. It seems likely that more examples will emerge. They will be of interest not only for what they teach us about the causes of such illnesses and possible therapies for them (for instance, the effectiveness of negative antagonists), but also for the clues they offer about the structural basis for regulation of the fundamental process of receptor activation. □

Robert J. Lefkowitz is in the Howard Hughes Medical Institute Research Laboratories, Duke University Medical Center, Durham, North Carolina 27710, USA.

ELECTRONICS

Chaos in harness

J. C. G. Lesurf

RESEARCH on chaos has sometimes been regarded as a flower amongst the vegetables — pretty to look at, but of limited practical use. Given the flood of computer graphics that have turned the 'Mandelbrot set' into an icon, this reaction is understandable. Despite this, many scientists and engineers have been exploring the potential of chaotic behaviour in applications far beyond the printing of colourful T-shirts (see the News and Views article by Abarbanel¹). Now Kevin Cuomo and Alan Oppenheim² are exploring how simple electronic circuits behaving like chaotic oscillators can be used in the fields of signal communications and data encryption.

Chaotic oscillations can occur in systems whose state variables have a nonlinear relationship; a good example is the pattern of atmospheric fluctuations whose behaviour has been called 'Lorenz oscillations' after their discoverer³. Chaotic oscillations vary in a way that can appear unpredictable. Despite being driven by a deterministic process, their fluctuations look like random noise.

Randomly varying quantities have a number of applications in information processing and communications. The best known is their use to mask or encrypt information where they offer two advantages to anyone who wishes to send secret messages. First, signals randomized in this manner may pass unnoticed against

the general background of natural noise. Second, if detected, their unpredictable variations give no clue to the information they contain. Of course, encoding information to look random must be done in a way that the intended recipient can unscramble.

Over the years many spies have produced secret messages using a series of random numbers printed on a 'one time pad'. Provided the numbers on the pad are a truly random sequence (and used only once), the encrypted message has all the statistical properties of random noise. As a result, it is impossible to decipher without a copy of the pad. Modern communications systems replace the pad with a digital computer which calculates a series of apparently random (or 'pseudo-random') numbers. The key to decoding the message is to know the details of how these numbers were computed. Although much faster than pencil and paper, this approach is limited by the computational speed of a conveniently sized and priced computer. Problems can also arise because of the way conventional digital computers store numbers as finite series of digits.

An excellent way to overcome these problems is to replace the digital computer with an analogue circuit. Systems like this, containing just a few components, have the potential to generate random sequences at rates equivalent to billions

1. Parma, J. *et al.* *Nature* **365**, 649–651 (1993).
2. Shenker, A. *et al.* *Nature* **365**, 652–654 (1993).
3. Rosenthal, W., Antaramian, A., Gilbert, S. & Birnbaumer, M. *J. biol. Chem.* **268**, 13030–13033 (1993).
4. Tsigos, C., Arai, K., Hung, W. & Chrousos, G. P. *J. clin. Invest.* (in the press).
5. Cotecchia, S., Exum, S., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2896–2900 (1990).
6. Lefkowitz, R. J., Cotecchia, S., Samama, P. & Costa, T. *Trends pharmac. Sci.* **14**, 303–308 (1993).
7. Okamoto, T. *et al.* *Cell* **67**, 1–20 (1991).
8. Cheung, A. H., Huang, R. R. C., Graziano, M. P. & Strader, C. D. *FEBS Lett.* **279**, 277–280 (1991).
9. Kjelsberg, M. A., Cotecchia, S., Ostrowski, J., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **267**, 1430–1433 (1992).
10. Samama, P., Cotecchia, S., Costa, T. & Lefkowitz, R. J. *J. biol. Chem.* **268**, 4625–4636 (1993).
11. Ren, Q., Kurose, H., Lefkowitz, R. J. & Cotecchia, S. *J. biol. Chem.* **268**, 16483–16487 (1993).
12. Robbins, L. S. *et al.* *Cell* **72**, 827–834 (1993).
13. Robinson, P. R., Cohen, G. B., Zhukovsky, E. & Oprian, D. D. *Neuron* **9**, 719–725 (1992).
14. Julius, D., Livelli, T. J., Jessell, T. M. & Axel, R. *Science* **244**, 1057–1062 (1989).
15. Allen, L. E., Lefkowitz, R. J., Caron, M. G. & Cotecchia, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11354–11358 (1991).

RÉSUMÉ

of numbers per second⁴. They make use of chaotic oscillations produced by nonlinear feedback. As their state is determined by smoothly variable quantities, an analogue system can avoid the rounding and finite-step errors that arise in digital computation.

In principle, a secure communications system could use two identical chaotic oscillators, one in the transmitter and one in the receiver. The transmitter oscillator's noise-like output is used to mask or scramble the message. The receiver's oscillator would then separate out the message from the artificial noise and recover the secret information. Unfortunately, one of the fundamental properties of any chaotic system is a sensitivity to initial conditions. Any infinitesimal differences between two nominally identical systems will cause them to produce very different output patterns, and it is impossible to make two electronic circuits that are absolutely identical.

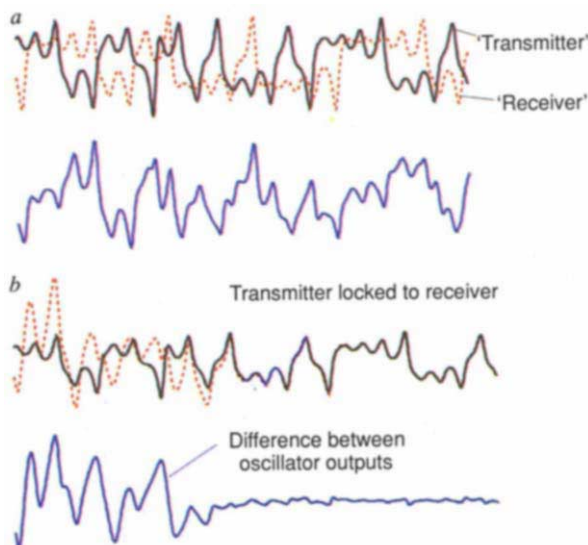
Various ways around this problem have been tried. For example, some work has been done on semi-chaotic oscillators. These produce signals that initially appear chaotic, but repeat themselves after a long interval. They have some of the properties of conventional linear oscillators and it is possible to make pairs similar enough to use in a transmitter-receiver arrangement.

A more powerful alternative technique is to 'lock' two similar chaotic oscillators together. Despite the interest in this possibility, the published work has largely been a theoretical examination of mathematical possibilities⁵⁻⁷. Cuomo and Oppenheim, in contrast, offer a demonstration of how a specific analogue circuit can be used in practice. The technique they use is similar to 'injection locking' which engineers routinely employ to synchronize conventional periodic oscillators. The notable feature of their report is that they were able to communicate information using a pair of synchronized chaotic oscillators. They describe two simple methods.

First, a digital signal was used to modulate one of the parameters of the transmitting oscillator. When a zero was being transmitted the transmission oscillator was nominally identical with the receiver's. When a one was being sent the transmission oscillator's behaviour was altered so much that the receiver was unable to produce a similar output. The pattern of received data could then be

recognized from noting when the receiver's oscillator was able to shadow the received signal (0) and when it was not (1).

Their second method used the chaotic signal to mask a speech pattern. The two were added together at the receiver, using a speech signal power 20 to 40 decibels below the chaotic signal level, to form a combination that looked like random noise with no obvious speech features. The receiver oscillator's output was then



It is possible to synchronize the apparently random variations of two chaotic oscillators by injection locking: the received signal is injected into the receiver oscillator at a point in the circuit equivalent to where it was taken from the transmitter oscillator. *a*, Typical output patterns that might be obtained from a pair of nominally identical oscillators, one in a transmitter, the other in a receiver, and the difference between the signals. *b*, The receiver oscillator locked to the transmitter. After a transient struggle, the receiver oscillator's output is pulled into synchronism with the transmitter's.

subtracted from the transmitted signal to reveal the speech pattern. Although the recovered pattern was not identical to the initial speech, it was, say the authors, of reasonable quality.

The examples given established that a receiver containing the correct circuit can recover chaotically encrypted information. To any receiver without the necessary oscillator the transmitted signal should be indistinguishable from noise.

Perhaps in future we should look more carefully at T-shirts decorated with Mandelbrot pictures. There may be a secret message hidden in the pattern. □

J. C. G. Lesurf is in the Physics and Astronomy Department, University of Saint Andrews, Fife KY16 9SS, UK.

1. Abarbanel, H. D. I. *Nature* **364**, 672–673 (1993).
2. Cuomo, K. M. & Oppenheim, A. V. *Phys. Rev. Lett.* **71**, 65–68 (1993).
3. Lorentz, E. N. *Atmos. Sci.* **20**, 130 (1963).
4. Robertson, M. R. & Lesurf, J. C. G. *Int. J. Infrared Millimetre Waves* **12**, 1379 (1991).
5. Bak, P. *et al. Phys. Scripta* **T9**, 50 (1985).
6. Pecora, L. M. & Carroll, T. L. *Phys. Rev. Lett.* **64**, 821 (1990).
7. Pecora, L. M. & Carroll, T. L. *Phys. Rev.* **A44**, 2374 (1991).

One man's meteorology. . .

THE global positioning system (GPS) of satellites and ground-based or low-Earth-orbit receivers was intended as a navigational tool, but it is also popular with geophysicists and geodesists because of the positional information it provides. To a geodesist, the effects of the Earth's atmosphere on the travel times of the radio waves are a nuisance. L. L. Yuan and co-workers, being atmospheric scientists, see it rather differently. To them, this noise is replete with information about the atmosphere (*J. geophys. Res.* **98**, 14925–14937; 1993). Using simulated atmospheric data from an established global climate model, they show that two chief effects of doubling carbon dioxide in the atmosphere (raised temperatures and greater water vapour content) could be detected without putting any new equipment into space.

Cross connection

BRITISH badgers have been gassed and trapped since they were implicated in the transmission of tuberculosis to cattle. Levels of bovine TB remain stubbornly high in parts of the south-west of the country, and P. White, J. A. Brown and S. Harris set out to find out why (*Proc. R. Soc. B* **253**, 277–284; 1993). Ingestion of grass contaminated with badger urine is one likely form of disease transmission to cattle. White *et al.* find that badgers tend to urinate on pasture after crossing a linear feature (such as sheep netting or a hedgerow), and that linear features tend to be commoner in areas where badgers have been subject to repeated control measures because of the high incidence of bovine TB. So enticing is the resulting hypothesis that one hopes it holds water.

In a twist

ATTEMPTS to relate the stability of α -helices to their constituent amino acids have relied mainly on empirical correlations. But J. M. Scholz *et al.* (*Biochemistry* **32**, 9668–9676; 1993) have now found a way to evaluate the contributions of side-chain interactions. Into a neutral water-soluble helix of alanine and glutamine residues they have put a pair of oppositely charged side-chains, one glutamate and one lysine, at different spacings. When the charges are separated by two or three host amino acids the helix is stabilized, but when they are next or next-but-one to each other they do the helix no good. The favourable interaction when the spacing is right is apparently dominated by hydrogen bonding. The glutamate, however, interacts just as favourably with a glutamine, but only when they are the right way round and separated by three residues, so in this case the steric demands for a good hydrogen bond are more restrictive. No doubt theorists will now come forward and predict that it must all be just so.

Fatty feedback and fluidity

Bruno Maresca and Andrew R. Cossins

How do eukaryotic cells maintain the physical structure of their membrane lipid bilayers within tolerable limits? The short answer is that fatty acid unsaturation and cholesterol contents are varied to provide the appropriate lipid milieu for membrane function¹. Yet little is known about how this is achieved and absolutely nothing about the cellular sensor concerned. A clue to what is involved now comes from Vigh, Murata and colleagues² — they show that the catalytic hydrogenation of a small pool of plasma membrane fatty acids activates transcription of the gene for the enzyme concerned, which leads to the resynthesis of unsaturated fatty acids.

The authors have used a simple, well-defined system, a strain of the blue-green alga *Synechocystis*, PCC6803, which adaptively regulates membrane lipid unsaturation when growth temperature is altered. Cooling causes the incorporation of increased proportions of unsaturated fatty acids, and this effectively fluidizes the bilayer and offsets the rigidity induced by cold. Murata *et al.*³ previously cloned the *desA* gene which encodes the Δ^{12} -desaturase; this is the enzyme that inserts an unsaturation bond into the

saturated fatty acid and thereby modifies its physical properties. They have also shown that lowering the growth temperature leads to a marked increase in the level of *desA* transcription⁴.

Temperature, however, is a blunt instrument for dissecting the mechanisms involved in this response, because it is totally nonspecific in its effects on cellular metabolism. So Vigh *et al.* employed the technique of homogeneous catalytic hydrogenation⁵, which allows the controlled *in situ* hydrogenation of unsaturated membrane lipids with apparently no effects other than the ordering of the membrane interior. The latest work² shows that hydrogenation of *Synechocystis* stimulates the transcription of the *desA* gene in the same way as cooling. Because only the plasma membrane is hydrogenated, this membrane must act as the sensor to altered saturation or 'fluidity', thereby initiating a control response, the ultimate effect of which is increased transcription of *desA*. Moreover, hydrogenated cells display a heightened sensitivity to cooling in the activation of *desA* transcription, indicating that hydrogenation and cooling have additive effects and

that cooling has its effects through changes in membrane fluidity. The results offer direct evidence that membrane structure affects gene transcriptional activity, and point to the control of membrane lipid biosynthesis by a negative feedback loop based on lipid order.

These insights come at an exciting time, for the transcription of eukaryotic desaturase genes is now known to be regulated by other environmental factors. Transcript levels in yeast are influenced by the supplementation of culture media with unsaturated fatty acids^{6,7}, and similar effects have been observed in cultured murine lymphocytes supplemented with arachidonic acid⁸. The transcriptional regulatory sensor in yeast has a high degree of specificity for Δ^9 -fatty acids and is presumably mediated by the specific binding of the Δ^9 -fatty acid to a regulatory protein. But although this allows regulation of specific levels of free fatty acid, it does not explain induced gene transcription following membrane hydrogenation.

Meanwhile, the regulatory elements that control the transcription of desaturase genes are becoming clearer. A *cis*-acting DNA sequence element in the promoter of the desaturase gene has been shown to repress a reporter gene in pre-adipocytes but not in adipocytes. This element is located between -435 and -410 nucleotides upstream from the gene and binds a 58K protein, a putative repressor whose activity decreases during pre-adipocyte differentiation^{9,10}. Oleic acid-responsive elements have also been identified in genes coding for peroxisomal proteins of the yeast *Candida tropicalis*¹¹. These consensus sequences may be widespread, and may control a variety of genes during cell differentiation and hormonally induced lipid responses in higher animals, as well as during thermal and dietary disturbances in other creatures. Taken together these findings offer hints as to the control systems that are probably involved, although exactly how they are

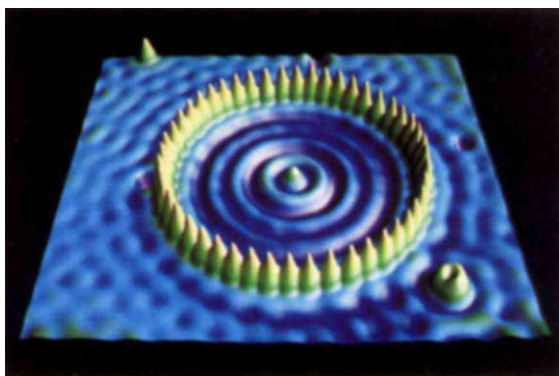
Round 'em up

QUANTUM confinement takes on a new meaning in experiments described last week by Don Eigler and colleagues (*Science* 262, 218–220; 1993). They have built an atomic-scale corral from 48 iron atoms deposited on a copper surface, within which the electrons of the copper's surface states are trapped.

When confined on this scale, electrons reveal their wave-like nature. In the 'quantum corral' staked out by Eigler and colleagues, the electron waves take the form of concentric ripples in electron density — standing waves set up by the scattering of electrons by the iron atoms. Oscillations in the density of states of surface electrons in a metal can be imaged using the scanning tunnelling microscope (*Nature* 363, 524; 1993), as the tunnelling current reflects this density of states. In the latest striking demonstration of how the STM can be used as a versatile tool for nanotechnology, the researchers first use the microscope tip to drag the iron atoms into a circle (see *Nature* 344, 524; 1990) and then switch to imaging mode to study the electrons confined inside.

The behaviour of electrons in very small confined systems (quantum 'dots' and 'wires') is attracting great interest both because of the new physics that such systems exhibit (for example, quantized charging effects, ballistic motion and chaotic behaviour) and because of the potential technological applications of nanoscale electronic devices. The demonstration that such structures can be built atom-by-atom with essentially arbitrary geometries, and that their electronic properties can be mapped out in real space, is sure to be greeted with an intense flurry of activity.

Philip Ball



1. Quinn, P. J., Joo, F. & Vigh, L. *Prog. Biophys. molec. Biol.* **53**, 71–103 (1989).
2. Vigh, L., Los, D. A., Horvath, I. & Murata, N. *Proc. natn. Acad. Sci. U.S.A.* **90**, 9090–9094 (1993).
3. Wada, H., Gombos, Z. & Murata, N. *Nature* **347**, 200–203 (1990).
4. Los, D., Horvath, I., Vigh, L. & Murata, N. *FEBS Lett.* **318**, 57–60 (1993).
5. Joo, F. *et al.* *Analyt. Biochem.* **194**, 34–40 (1991).
6. Bossie, M. A. & Martin, C. E. *J. Bact.* **171**, 6409 (1989).
7. McDonough, V. M., Stuke, J. E. & Martin, C. E. *J. Biol. Chem.* **267**, 5931–5936 (1992).
8. Tebbey, P. W. & Buttke, T. M. *Biochim. biophys. Acta* **1171**, 27–34 (1992).
9. Swick, A. G. & Lane, D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7895–7899 (1992).
10. Vasseur-Cognet, M. & Lane, M. D. *Curr. Opin. Genet. Dev.* **3**, 238–245 (1993).
11. Sloots, J. A. *et al.* *Gene* **105**, 129–134 (1991).
12. Wada, H. & Murata, N. *Pl. Cell Physiol., Tokyo* **30**, 971–978 (1989).
13. Stewart, L. C. & Yaffe, M. P. *J. Cell Biol.* **115**, 1249–1257 (1991).
14. Wolter, F. P., Schmidt, R. & Heinz, E. *EMBO J.* **11**, 4685–4692 (1992).
15. Murata, N. *et al.* *Nature* **356**, 710–713 (1992).
16. Szidonya, J., Farkas, T. & Pali, T. *Biochem. Genet.* **28**, 233–246 (1990).

influenced by the membrane's physical condition remains obscure.

However, the new approaches do allow a much more direct investigation of the functional importance of membrane unsaturation and physical structure at the whole organism level, as well as that of the cell. Murata's group has shown that disruption of the desaturase gene interferes with the normal chill tolerance of cyanobacteria¹² and they have manipulated the chill tolerance of a chill-sensitive species by transfection with *desA* (ref. 3). In *Saccharomyces*, a temperature-sensitive mutation has been observed within the Δ^9 -desaturase gene and temperature-sensitive growth and defective mitochondrial movements are complemented by supplementation of cultures with oleic acid¹³. In higher plants, such as *Arabidopsis* and tobacco, genetic manipulation of lipid saturation causes an

increased chill tolerance^{14,15}. And in *Drosophila* a temperature-sensitive mutant is defective in the regulation of lipid unsaturation in the cold, possibly by disruption of desaturase activity¹⁶.

The combination of catalytic hydrogenation with these powerful molecular techniques promises a great deal. It will add momentum to attempts to discover how lipid composition is controlled, to manipulate the thermotolerance of commercially important organisms, and to reveal the signal transduction pathways involved in the adaptive regulation of gene expression. □

Bruno Maresca is at the International Institute of Genetics and Biophysics, Via Marconi 12, 80125 Naples, Italy. Andrew R. Cossins is in the Department of Environmental and Evolutionary Biology, University of Liverpool, Liverpool L69 3BX, UK.

X-RAY ASTRONOMY

Interloper apprehended

Jules P. Halpern

UNTIL recently, the origin of the apparent 3.4-hour periodicity in the X-ray emission from the Seyfert galaxy NGC6814 was a notable astronomical puzzle. Among active galactic nuclei (AGNs), it was the only strictly periodic signal reported at any wavelength that passed tests of stability and persistence¹. Exotic theories contrived to explain this unique phenomenon left many researchers uneasy. These proposals included gravitational lensing of orbiting hotspots or vortices in an accretion disk, and occultation by an orbiting star captured by the black hole thought to be at the centre of all AGNs. Now, on page 626 of this issue, Madejski *et al.*² have ended the mystery by showing that, although the period is indeed real, it belongs not to NGC6814 but to another X-ray source some 37 arcminutes away in the sky.

The interloper is a new cataclysmic variable, almost certainly of the AM Herculis type in which a highly magnetized white dwarf accretes matter donated from a main-sequence companion star (about 20 such stars are known). The 3.4-hour modulation represents the orbital period of the binary.

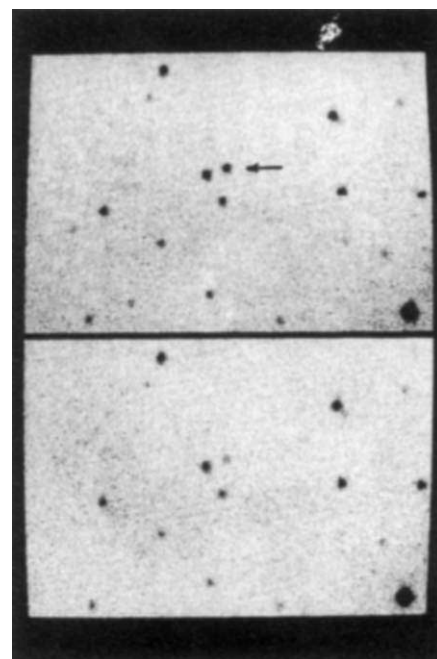
With the benefit of hindsight, it is easy to chastise observers for not noticing that the X-ray period was detected only in observations of NGC6814 using non-imaging detectors, which had fields of view of more than 1 degree. If the newly identified X-ray source had not fallen just outside an X-ray image of NGC6814 obtained by the Einstein satellite 14 years ago, this mistake would probably never have been made. Only now, with the

wide-field imaging capability of the Rosat satellite, has the culprit been apprehended.

Soon after the news broke, optical photometry of the field was obtained with the 32-cm robotic telescope and CCD camera at the Center for Basement Astrophysics³ in the suburbs of Washington DC. Sure enough, there is a variable star at the X-ray position, which eclipses with a strict period of 3.4 hours⁴.

This is certainly the happiest possible end to a prolonged episode of observational bad luck that for three years fuelled the imaginations of observers and theorists. It is perhaps a comment on our primitive understanding of AGNs that theories for a Seyfert galaxy can be shoehorned into data from a star that is millions of times smaller. The evidence for periodic signals from AGNs has always been unsatisfying, and is now virtually nonexistent. One can't help wondering how reliable is the claim of quasi-periodic X-ray oscillations from another Seyfert galaxy, NGC5548 (ref. 5), which has yet to be confirmed, and which is even more difficult to understand theoretically⁶, because the periods are so short — only 500–1,000 seconds.

In the case of NGC6814, it is important to realize that not just the X-ray period, but all of the X-ray properties previously attributed to it have to be abandoned. Some of our more interesting ideas about Seyfert galaxies were motivated by the additional, non-periodic variations of this supposedly unique X-ray source. Most notable was the rapid and simultaneous variability of the continuum and the iron



Now you see it, now you don't: CCD images from the Center for Basement Astrophysics³ showing the variable-star counterpart of RXJ1940.2–1025 near maximum and minimum light, at magnitudes 16.0 and 16.9, respectively. The field shown is 4 × 2 arc minutes, and is stretched slightly in the north–south direction. North is up, east is left.

K α emission line⁷, which had been interpreted as reprocessing of continuum X-rays into fluorescent emission on the inner accretion disk, but is really just the conventional behaviour of an AM Her star. The data fuelled expectations that high-resolution X-ray spectroscopy of Seyfert galaxies would reveal large rotational velocities in an accretion disk within 50 Schwarzschild radii of the black hole. This hope must now be dampened, as there is no longer any evidence for rapid emission-line variability, and hence for such a compact line-emitting region, in Seyfert galaxies. Even before the discovery of the period, NGC6814 was thought to be the most rapidly variable Seyfert galaxy, implying an unusually small black hole mass (for an AGN) of about 10⁶ solar masses⁸. It will have to relinquish this role as well.

Although NGC6814 is now relegated to comparative obscurity, the new X-ray source, designated RXJ1940.2–1025, will surely become one of the most celebrated and best studied AM Her stars. In this class of cataclysmic binary, which are also called 'polars', the magnetic field of the accreting white dwarf is strong enough

1. Done, C. *et al.* *Astrophys. J.* **400**, 138–152 (1992).
2. Madejski, G. M. *et al.* *Nature* **365**, 626–628 (1993).
3. Skillman, D. R. *Sky Tel.* **85**, 83–84 (1993).
4. Patterson, J., Halpern, J. & Skillman, D. R. *IAU Circ. No.* 5852 (1993).
5. Papadakis, I. E. & Lawrence, A. *Nature* **361**, 233–236 (1993).
6. Halpern, J. P. *Nature* **361**, 203–204 (1993).
7. Kunieda, H. *et al.* *Nature* **345**, 786–788 (1990).
8. Wandel, A. & Mushotzky, R. F. *Astrophys. J.* **306**, L61–L65 (1986).
9. Cropper, M. *Space Sci. Rev.* **54**, 195–295 (1990).

to channel material from the main-sequence companion directly onto one of its magnetic poles, without the formation of an accretion disk. The white dwarf is locked in synchronous rotation with the orbit, causing strong modulation of the X-ray and optical light at the orbital period. Both the X-ray and optical light curves, originating in the accretion column, are characteristic of AM Her stars⁹, as is the 'two-component' X-ray spectrum². It is a safe wager that the star will show strong emission from ionized helium (He II), circular and linear polarization, and the usual panoply of AM Her traits. There is now a long historical record of 'serendipitous' X-ray eclipses

which can be used to study the orbital period evolution. Another fortunate circumstance is that the spectrum indicates similar observed fluxes in 'hard' and 'soft' X-ray components², implying that the relationship between these components can be directly investigated. This relationship has been the subject of much controversy⁹, and there are precious few objects like RX J1940.2-1025 whose hard X-ray flux is high enough to make a detailed study feasible. □

Jules P. Halpern is in the Columbia Astrophysics Laboratory, Columbia University, 538 West 120th Street, New York, New York 10027, USA.

GLACIOLOGY

Stopping an ice stream

Gabrielle Walker

AN ice stream in the Antarctic has stopped moving, and glaciologists are scrambling to find out why. Given how large the polar ice caps loom in possible global responses to climate change, climatologists are also eagerly awaiting a resolution of this question, which was a topic of lively discussion at a meeting last month*.

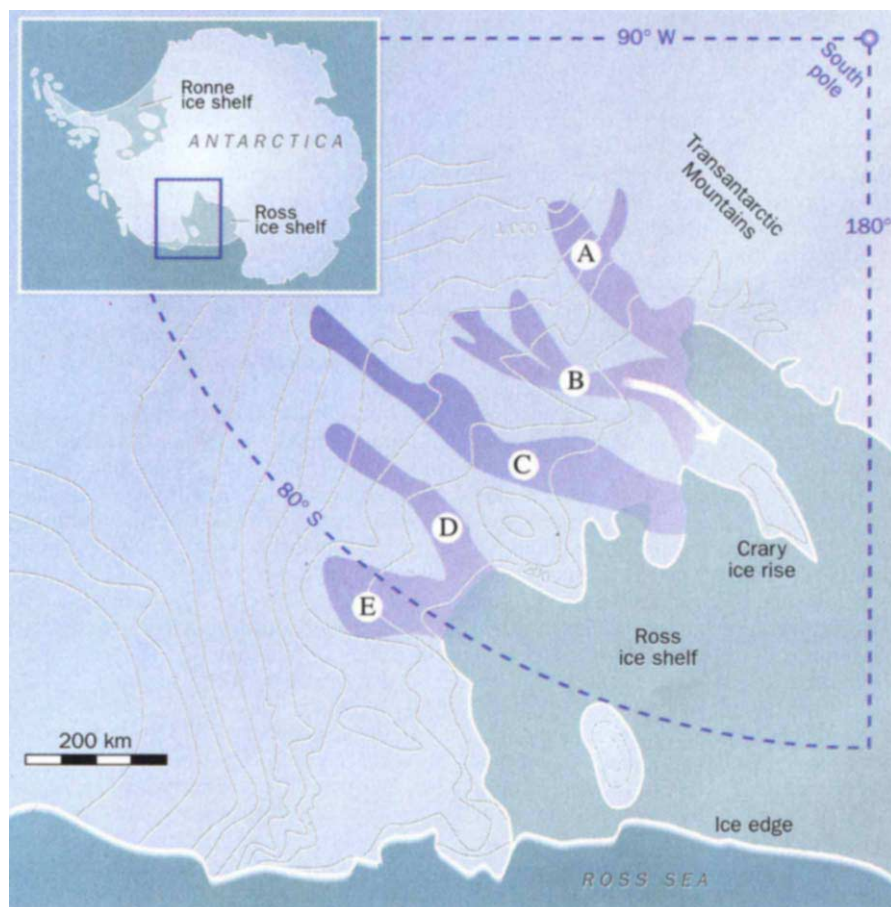
The ice stream concerned lies in the West Antarctic Ice Sheet, itself of especial interest because it is thought to be the most vulnerable of such sheets to decay. Its size is regulated by the accumulation of snow in the interior and loss of ice as bergs, which break off into the surrounding sea. Much of that ice is transported from the interior by fast-moving ice streams and then tipped out onto the ocean surface to form part of one of two vast floating ice shelves — the Ross and the Ronne shelves. But ice stream C, one of the Ross ice streams (see map), has come to a halt, and new evidence suggests that this happened very recently in geological terms (and not at a time of a dramatic shift in climate). Apart from the issues of why and how stream C has stopped, there is that of whether the others might follow suit. If they do, the interior ice would take much longer to drain to the ice shelves, and the sheet could become thicker and more stable, locking up sea water as snow and ice and lowering global sea level.

Ice stream C is now covered with snow, but radar images show the same flow lines and crevasses that characterize the surfaces of the neighbouring, active streams. Retzlaff and Bentley¹ used the depth of the snow covering the crevasses at various points along the edge of the stream to determine how long ago the ice stopped

moving (as crevasses are continually manufactured at the edge of a moving ice stream, their sudden disappearance provides a useful 'time marker' of ice-stream stoppage). They estimate that the lower, wide part came to a virtual halt around 130 years ago, whereas the upper part still

seems to be moving, although it is not yet clear how fast.

At the meeting, Richard Alley (Pennsylvania State University) and colleagues, Bentley among them, proposed that the key to the braking mechanism could lie with events at the base of the ice streams². Stream B, the best studied of the five Ross ice streams, is known to move by sliding over a layer of soft, well-lubricated, almost frictionless till. From the flow patterns revealed by the radar images, it seems that stream C once moved at least as quickly as its neighbour, so it is reasonable to assume that stream C too had a layer of wet deformable till at its base. The initiation point of stream B has begun to move back towards the interior of the ice sheet³, and Alley and colleagues suggested that this is because of new lubrication caused by increased basal temperatures (the base of the ice stream is only now experiencing the increase in global temperature that occurred at the surface more than 10,000 years ago when the Earth moved out of the last ice age⁴). But they also noted that similar movement of the main portion of stream C in the past few centuries would have taken its initiation point to a region with pronounced basal topography, which could have influenced the direction in which water



Antarctic ice streams (A-E) feeding ice through to the Ross ice shelf (adapted from I. M. Whillans and C. J. van der Veen, *J. Glaciol.* (in the press)).

*Fifth International Symposium on Antarctic Glaciology, British Antarctic Survey, Cambridge, UK, 5-10 September 1993.

flows at the base of the stream.

Usually, the basal water flow is determined by the slope of the ice surface, rather than that of the bed. But if, as Alley and colleagues suggest, the relatively flat ice stream extended inland over the steep bedrock, the water needed to lubricate the ice stream's till may have been forced down the bed slope away from stream C and towards stream B. So stream C's till would have lost its lubrication, and stream B would have 'stolen' the water. This 'water piracy' principle could also explain why stream B seems to be speeding up.

Sridhar Anandakrishnan (Pennsylvania State University) pointed out that to dehydrate the till sufficiently to stop the ice stream from moving would take around a million years⁵. However, he also suggested that 'sticky spots' in the stream bed could provide the solution. He reported evidence for microearthquakes at the base of ice stream C, occurring at discrete zones of high friction between the ice and the underlying bedrock. These 'sticky spots' could be caused by islands of bedrock sticking up through the till layer, or by locally thin basal till. Either way, water piracy could have been enough to dewater these regions and bring the ice stream to a halt within the time required.

Retzlaff and Bentley (University of Wisconsin-Madison) have a different theory¹. They suggest that stream C may have had too much water, not too little. The stable configuration for free water at the ice-bed interface is thought to be a relatively uniform sheet. But if the water layer grew too thick it could become unstable. Thicker parts could then grow at the expense of thinner ones, channels could form, and this would lead to a drop in water pressure. Water would then drain from the till, greatly increasing the basal drag on the ice stream.

One drawback of this theory is that it is not clear why the water layer beneath stream C should have become thick enough to initiate this sequence of events. Also, even if water channels were to form, it is possible that till would be sucked into them, blocking water flow and causing a new build-up of water pressure. On the other hand, the advantage of the theory is that it seems to fit the relative timing of the stoppage. According to Alley's theory, this should happen first at the head of the stream and then feed through to the lower regions. But as the free-water layer beneath ice streams is thicker towards the lower end, Retzlaff and Bentley's mechanism would mean that stagnation should start there. In fact, the snow thicknesses

above the crevasses seem to indicate that the stagnation began downstream, but Retzlaff and Bentley are careful to stress that the dates throughout the ice stream are indistinguishable within error limits.

The next step is to devise experiments to distinguish between the theories. If Alley and colleagues are right, the stoppage was simply a freak of the basal topography below ice stream C. If Retzlaff and Bentley are right, then in principle any or all of the ice streams might experience excessive basal water, and cease

moving. It is even possible that this might happen in response to global warming, potentially counteracting any rise in sea level due to thermal expansion of the ocean. But by the time any anthropogenic temperature rise could feed through to the base of the ice to cause melting and set the process in motion 10,000 years hence, we would all, presumably, have other things on our minds. □

Gabrielle Walker is an assistant editor at Nature.

BIODIVERSITY

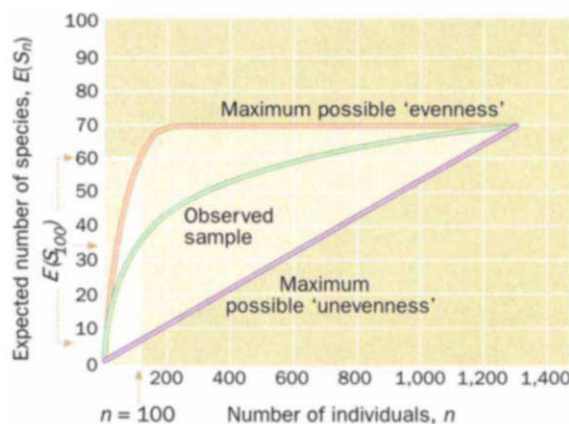
A dip into the deep seas

John D. Gage and Robert M. May

As understanding of life on the deep-sea bottom increases, ecological patterns are emerging which often parallel those found on land and in shallow water. Take, for example, the letter of Rex *et al.* on page 636 of this issue¹, which reveals a large-scale decrease in species richness among small invertebrates in deep-sea sediments

and South Atlantic are not much different from the forms of life in soft sediments in coastal seas. They are, however, generally much richer in species. Why there should be such exuberant biological diversity in an environment apparently lacking in the habitat complexity of, say, tropical rain forest — whose species richness it might rival³ — remains an enigma.

Like high species richness, the LSDG has been the subject of various explanations^{4,5}. They range from so-called 'equilibrium' theories related to time (where either the warm, humid tropics provide a more favourable milieu for rapid evolutionary diversification, or the length of time or stability in physical conditions allow the biota to speciate in an uninterrupted fashion), to theories related to small-scale environmental complexity (which, although the argument is often a bit circular, is more marked in the tropics). On the face of it, the continuity and homogeneity in physical conditions in the deep ocean would seem to rule out these sort of explanations. And the endless sameness of the deep-sea bed does not encourage ideas centred on habitat heterogeneity, or indeed on other suggested differences



The rarefaction method. The middle curve is a rarefaction of expected species diversity of benthic invertebrates in a large sample (57 species among 1,318 individual specimens) from a sea loch in Scotland. The lower line shows the rarefaction of a theoretical sample also composed of 57 species and 1,318 individuals, but this time with 1,318 – 56 (1,262) individuals of one species and the remainder as singletons (maximum possible 'unevenness'). The upper curve shows a rarefaction of a sample where each of the 57 species has an equal abundance (23) totalling 1,311 individuals, representing a sample of very similar total size with maximum possible 'evenness'. Comparison of the expected number of species in samples of 100 individuals, $E(S_{100})$, will clearly be affected by the degree to which they show similar distributions of abundances amongst the species.

between the Equator and the poles. This so-called 'latitudinal species-diversity gradient' (LSDG) is well established for terrestrial plants and animals; for instance, the number of species of ants steadily declines from 222 down to 3 along a line drawn from Brazil to Alaska².

The communities of deep-sea creatures that Rex *et al.* have sampled in the North

in competition, predation or productivity. More recent 'nonequilibrium' explanations are based on gradients in the intensity, frequency and physical scale of disturbances. Such ideas encourage a view that similar mechanisms operate in the terrestrial tropics and deep-sea benthos, which could promote coexistence of many species at small spatial scales. They are,

1. Retzlaff, R. & Bentley, C. R. *J. Glaciol.* (in the press).
2. Alley, R. B., Anandakrishnan, S., Bentley, C. R. & Lord, N. *Ann. Glaciol.* (in the press).
3. Shabtaie, S., Bentley, C. R., Bindshadler, R. A. & MacAyeal, D. R. *Ann. Glaciol.* **11**, 137–149 (1988).
4. Whillans, I. M. *J. geophys. Res.* **86**, 4274–4282 (1981).
5. Anandakrishnan, S. & Alley, R. B. *Ann. Glaciol.* (in the press).

however, less helpful with respect to LSDG in the deep sea.

Rex and colleagues' results are somewhat unexpected, because the evidence for a similar gradient in the animals of coastal sediments is still weak⁶. But they prompt us to look at another of the largest ecological gradients on this planet, that of serial replacement of bottom-dwelling species with increasing depth in the deep ocean⁷. This phenomenon is associated with an increase in species richness with depth, and is essentially like the pattern of increasing numbers of plant and animal species as one moves down from mountain tops to sea level. Apart from broad parallels between ascending mountains and going from the Equator to the poles, the causes for these elevational gradients remain contentious.

It is too much to expect the findings of Rex *et al.* to settle these matters. This is not the fault of their data, which are the most comprehensive yet. But because of the enormity of the task of making taxonomic assignments for so many species (many new to science), the groups of animals analysed by Rex *et al.* represent much less than half the total diversity present in their original, carefully collected samples. Arguably this does not reduce the value of the analysis, because there is little evidence for resource partitioning amongst the organisms in question; most seem to be unspecialized feeders on the ooze. So the constraints operating on one group may be expected to operate on others. The similar LSDGs shown by the three groups investigated — gastropods, bivalves and isopods — support this view.

Another issue stems from the difficulty of estimating diversity from samples taken from the ocean floor. The standard 'rarefaction' method used by Rex *et al.* is based on estimating the number of species, S , expected to be found in a sample of n individuals, $E(S_n)$. By assessing the functional dependence of $E(S_n)$ upon n , one can hope to estimate the true number of species that would be found, asymptotically, in very large samples ($n \rightarrow \infty$). Consequently marine biologists usually use the expected number of different species in a sample of, say, 100 individuals, $E(S_{100})$, as a measure of the relative richness of species in comparisons among different habitats. Rex *et al.* use $E(S_{100})$ in this way. But the functional dependence of $E(S_n)$ upon n , and thence $E(S_{100})$ in a particular community, obviously depends on the number of species actually present and on the relative abundance of the different species⁸. The figure illustrates these complications, which must be kept in mind as factors that could cloud interpretation of the patterns in LSDGs recorded by Rex *et al.*

A less well-recognized latitudinal gradient on land is one in which tropical

species seem to have smaller geographical ranges than species at higher latitudes^{9,10}. This phenomenon has recently been the subject of analysis for mammals and birds in the Nearctic^{11–13} (essentially North America) and Palearctic¹⁴ (essentially northern Europe), though doubts about its generality are raised by the absence of any such patterns among Australian mammals¹⁵. On land, one explanation for this narrowing of 'zonation' towards the Equator is that tropical species may have narrower environmental tolerances than those necessary at higher, harsher latitudes.

It would be nice to test these ideas in the deep sea, at increasing depths, using methods to describe the bathymetric patterns in species' ranges. Exploration of differences in species' ranges in ocean basins at different latitudes is likely to be tougher, partly because there has been less thorough sampling at low latitudes, and partly because accurate delineation of the range of a species on the deep-ocean bed is a herculean task. Such studies are likely to be further complicated by the varying standards that have been used in discussing ocean-floor 'species'; the new tools of molecular genetics could help here.

Maybe the deep-ocean floors are not so globally uniform after all. We now know that there is a seasonal, rapidly transmitted flux of organic particles from plankton blooms right down to the deep-sea bed at temperate¹⁶ and subpolar¹⁷ latitudes. If there is a polewards gradient in availability of this labile food, with associated habitat heterogeneity from the resulting patchy carpeting of detrital material on the sea bed, it may be a clue to maintenance of the LSDG. □

John D. Gage is at the Scottish Association for Marine Science, PO Box 3, Oban, UK. Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Rex, M. A. *et al.* *Nature* **365**, 636–639 (1993).
2. Fischer, A. G. *Evolution* **14**, 64–81 (1960).
3. Grassle, J. F. & Maciolek, N. J. *Am. Nat.* **139**, 313–341 (1992).
4. Begon, M., Harper, J. L. & Townsend, C. R. *Ecology: Individuals, Populations and Communities* (Blackwell, Oxford, 1986).
5. Rhode, K. *Oikos* **65**, 514–527 (1992).
6. Sanders, H. L. *Am. Nat.* **102**, 243–282 (1968).
7. Carney, M. A., Haedrich, R. L. & Rowe, G. T. in *Deep-Sea Biology* (ed. Rowe, G. T.) 371–398 (Wiley, New York, 1983).
8. May, R. M. *Nature* **361**, 598 (1993).
9. Simpson, G. G. *Syst. Zool.* **13**, 57–73 (1964).
10. Rapoport, E. H. *Areography* (Pergamon, Oxford, 1982).
11. Stevens, G. C. *Am. Nat.* **133**, 240–256 (1989).
12. Brown, J. H. & Maurer, B. A. *Science* **243**, 1145–1150 (1989).
13. Pagel, M., May, R. M. & Collie, A. R. *Am. Nat.* **137**, 791–815 (1991).
14. Letcher, A. J., Purvis, A., Nee, S. & Harvey, P. H. *J. anim. Ecol.* (in the press).
15. Smith, F. D. M., May, R. M. & Harvey, P. H. *J. anim. Ecol.* (in the press).
16. Billett, D. S. M., Lampitt, R. S., Rice, A. L. & Mantoura, R. F. C. *Nature* **302**, 520–522 (1983).
17. Graf, G. *Nature* **341**, 437–439 (1983).

DAEDALUS

High vacuum

THE balloon would be an ideal aircraft if you didn't have to fill it with gas. Hydrogen is inflammable, helium is expensive, and hot air needs to be replenished by burning fuel. All tie the craft down to ground-based supplies. So Daedalus proposes a gasless balloon. It is filled with vacuum, which is even lighter than hydrogen.

Sadly, no conventional balloon could stand being evacuated. External air pressure would put the envelope into strong compression, squashing it flat. So Daedalus is devising a cunning electromagnetic envelope which cancels this applied compression by a countervailing internal tension.

He points out that if two adjacent wires carry currents in opposite directions, they repel each other. Imagine, he says, a length of insulated wire, folded zigzag into a plane grid of many parallel wires. A current through the whole grid would traverse adjacent wires in opposite directions. Each wire would repel its neighbours. A second grid laid down at right-angles to the first would experience this internal repulsion in the perpendicular direction. Weave the wires together into a gauze, coat it with an impermeable film, and the result is Daedalus's novel balloon fabric. Energized by a current, it tends to expand in all directions in its own plane. It shows not surface tension but 'surface compression', which neatly cancels external compressive forces.

A single wire grid cannot be mapped onto a sphere. Like the lines of latitude and longitude, it would have at least two singularities or poles. So Daedalus's electromagnetic vacuum balloon will be tessellated from a number of separate gridded panels, overlapping at the joins. To tension a realistically sized vacuum balloon against full sea-level atmospheric pressure, its wires will have to carry a permanent and desperately heavy current; but Daedalus hopes that room-temperature superconductors will soon be along to solve that problem.

The resulting balloon will need no gas, and almost no fuel. On the ground, its envelope will be kept inflated with air. For a flight, the air will gradually be pumped out, while the tensioning current is built up in the superconducting wires of the envelope. Once established, the current will flow endlessly; the released balloon will sail noiselessly aloft. Its lift can be varied at will, simply by pumping air into or out of the envelope. It can rise and descend easily and repeatedly, without releasing sand-bags or venting gas, and can return to Earth anywhere at any time. It will be the most independent of flying machines.

David Jones

Music and spatial task performance

SIR—There are correlational¹, historical² and anecdotal³ relationships between music cognition and other 'higher brain functions', but no causal relationship has been demonstrated between music cognition and cognitions pertaining to abstract operations such as mathematical or spatial reasoning. We performed an

experiment in which students were each given three sets of standard IQ spatial reasoning tasks; each task was preceded by 10 minutes of (1) listening to Mozart's sonata for two pianos in D major, K488; (2) listening to a relaxation tape; or (3) silence. Performance was improved for those tasks immediately following the first condition compared to the second two.

Standard age scores for each of the three listening conditions.

Testing procedure. In the music condition, the subject listened to 10 min of the Mozart piece. The relaxation condition required the subject to listen to 10 min of relaxation instructions designed to lower blood pressure. The silence condition required the subject to sit in silence for 10 min. One of three abstract reasoning tests taken from the Stanford-Binet intelligence scale⁴ was given after each of the listening conditions. The abstract/spatial reasoning tasks consisted of a pattern analysis test, a multiple-choice matrices test and a multiple-choice paper-folding and cutting test. For our sample, these three tasks correlated at the 0.01 level of significance. We were thus able to treat them as equal measures of abstract reasoning ability.

Scoring. Raw scores were calculated by subtracting the number of items failed from the highest item number administered. These were then converted to SAS using the Stanford-Binet's SAS conversion table of normalized standard scores with a mean set at 50 and a standard deviation of 8. IQ equivalents were calculated by first multiplying each SAS by 3 (the number of subtests required by the Stanford-Binet for calculating IQs). We then used their area score conversion table, designed to have a mean of 100 and a standard deviation of 16, to obtain SAS IQ equivalents.

Thirty-six college students participated in all three listening conditions. Immediately following each listening condition, the student's spatial reasoning skills were tested using the Stanford-Binet intelligence scale⁴. The mean standard age scores (SAS) for the three listening conditions are shown in the figure. The music condition yielded a mean SAS of 57.56; the mean SAS for the relaxation condition was 54.61 and the mean score for the silent condition was 54.00. To assess the impact of these scores, we 'translated' them to

spatial IQ scores of 119, 111 and 110, respectively. Thus, the IQs of subjects participating in the music condition were 8–9 points above their IQ scores in the other two conditions. A one-factor (listening condition) repeated measures analysis of variance (ANOVA) performed on SAS revealed that subjects performed better on the abstract/spatial reasoning tests after listening to Mozart than after listening to either the relaxation tape or to nothing ($F_{2,35} = 7.08$; $P = 0.002$). The music condition differed significantly from both the relaxation and the silence conditions (Scheffe's $t = 3.41$, $P = 0.002$; $t = 3.67$, $P = 0.0008$, two-tailed, respectively). The relaxation and silence conditions did not differ ($t = 0.795$; $P = 0.432$, two-tailed). Pulse rates were taken before and after each listening condition. A two-factor (listening condition and time of pulse measure) repeated measures ANOVA revealed no interaction or main effects for pulse, thereby excluding arousal as an obvious cause. We found no order effects for either condition presentation or task, nor any experimenter effect.

The enhancing effect of the music condition is temporal, and does not extend beyond the 10–15-minute period during which subjects were engaged in each spatial task. Inclusion of a delay period (as a variable) between the music listening condition and the testing period would allow us quantitatively to determine the presence of a decay constant. It would also be interesting to vary the listening time to optimize the enhancing effect, and to examine whether other measures of general intelligence (verbal reasoning, quantitative reasoning and short-term memory) would be similarly facilitated. Because we used only one musical sample of one composer, various other compositions and musical styles

should also be examined. We predict that music lacking complexity or which is repetitive may interfere with, rather than enhance, abstract reasoning. Also, as musicians may process music in a different way from non-musicians, it would be interesting to compare these two groups.

Frances H. Rauscher

Gordon L. Shaw*

Katherine N. Ky

Center for the Neurobiology of Learning and Memory,
University of California,
Irvine, California 92717, USA

* Also at Department of Physics.

1. Hassler, M., Birbaumer, N. & Feil, A. *Psychol. Music* **13**, 99–113 (1985).
2. Allman, G. J. *Greek Geometry from Thales to Euclid* p. 23 (Arno, New York, 1976).
3. Cranberg, L. D. & Albert, M. L. in *The Exceptional Brain* (eds Obler, L. K. & Fein, D.) 156 (Guilford, New York, 1988).
4. Thorndike, R. L., Hagen, E. P. & Sattler, J. M. *The Stanford-Binet Scale of Intelligence* (Riverside, Chicago, 1986).

MyoD and c-fos expression

SIR—Trousche *et al.*¹ have reported that the down regulation of *c-fos* expression during muscle cell differentiation may result from the binding of the helix-loop-helix (HLH) proteins to a CANNTG motif, or E-box, that occurs within the *c-fos* serum response element (SRE), thereby excluding the binding of the serum response factor (SRF) to this element. We investigated the interaction of HLH proteins with the *c-fos* SRE when molecular clones for E12 were first isolated by screening a phage expression library with the SRE probe². We estimated that the dissociation constant for the myogenin/E12-SRE complex was 10^{-8} – 10^{-9} M⁻¹ by comparing the relative effectiveness of E-box elements from different genes to compete for binding in the electrophoretic mobility-shift assays. The relatively low affinity of myogenin/E12 for the SRE could result from differences in nucleotide identities at the internal dinucleotide and flanking sequences between the *c-fos* E-box and the consensus HLH binding site^{3,4}. By contrast, the dissociation constants for the SRE-SRF or the SRE-SRF/p62^{TCF} complexes have been reported to be 5×10^{-10} and $\leq 10^{-11}$ M⁻¹, respectively⁵. Collectively, these data indicate that it is unlikely that HLH proteins alone can significantly compete with SRF for binding to DNA *in vivo*.

It has been reported in other studies that the *c-fos* SRE is either equally active in muscle and non-muscle cells⁶, or that it activates muscle-specific expression⁷ when situated upstream from a minimal promoter. Further, a comparison of the

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

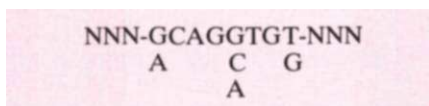
expression properties of a series of natural and synthetic SREs found no correlation between the presence of the E-box motif and activity in myocytes⁸.

SRF and SRF-related proteins control differentiation-specific and immediate-early gene expression by binding specifically to structurally similar elements occurring in the promoters of genes that respond to diverse regulatory controls. It seems that the complex array of regulatory functions conferred by these elements are mediated by the interactions of additional transcription factors including Phox1, YY1 and *ets*-related proteins⁹. The evidence for *c-fos* gene regulation by HLH proteins is intriguing, but further investigation is needed before it can be considered definitive.

Kenneth Walsh

Department of Physiology and Biophysics,
Case Western Reserve University,
Cleveland, Ohio 44106, USA

HAREL-BELLAN REPLIES — We maintain our belief that heterodimers formed by MyoD and E12 (or myogenin and E12) bind to the *c-fos* E-box with a high affinity¹. Indeed, *c-fos* SRE was compared to the canonical κ -E2-like E-box from the MCK promoter⁴ in competition experiments. Our results (see figure) indicate without ambiguity that the affinities of myogenic heterodimers for E-boxes from *c-fos* and MCK promoters do not differ significantly. Thus, the difference in sequence noted by Walsh does not dramatically impair the affinity of the myogenic proteins. In fact, this is not an unexpected result: the *c-fos* E-box (GCAGATGT in the reverse orientation) falls in the consensus binding sites for MyoD/E12 determined by Sun and Baltimore⁴:

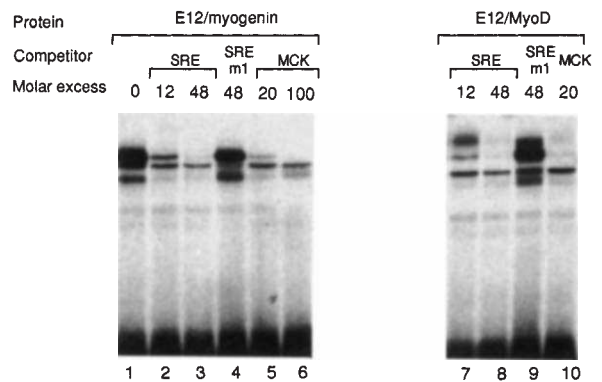


These authors estimated the dissociation constant for MyoD/E12 interaction with a κ -E2 site to be around $6 \times 10^{-11} \text{ M}^{-1}$. Thus, a dissociation constant for *c-fos* SRE of the order of 10^{-10} – 10^{-11} M^{-1} can be inferred from the results shown in the figure. This affinity may or may not be sufficient for MyoD/E12 to compete with SRF (and/or SRF/p62) *in vivo*, but it is clearly in a range in which this competition is possible. Note that competition is not the only likely hypothesis for inhibition

and we did suggest alternative mechanisms in our paper¹.

The finding that *c-fos* SRE has a similar activity in muscle and non-muscle cells⁶, independent of the presence of the E-box⁸, does not contradict our results. Indeed, we did not demonstrate a tissue-specific extinction of SRE activity. The similar activity rather suggests that, in muscle-determined cells (myoblasts), the SRE activity depends on the status of the cell with regard to proliferation: SRE activity seems to be blocked only when the cell enters a terminal differentiation process. In that regard, it is of primary importance that the test cells represent a homogeneous and synchronized population.

Walsh and colleagues do not mention a treatment of test cells which would induce terminal differentiation in their papers^{6,8}, so normal *c-fos* expression and SRE activity are expected. In addition, in the paper by Tuil *et al.*⁷ cited by Walsh as evidence for a muscle-specific activity of the SRE, the oligonucleotide referred to as SRE did not



EMSA analysis of E12/myogenin and E12/MyoD heterodimers. comparison between E-boxes from *c-fos* SRE and MCK promoter. *In vitro* translated proteins (as indicated) were analysed using an MCK canonical E-box as a probe, in the presence of indicated molar excess of either an SRE oligonucleotide or an MCK oligonucleotide, as indicated.

METHODS. Plasmid pMSV myogenin (a gift from E. Olson), pMSV MyoD and pMSV E12 were transcribed *in vitro* and translated as described¹. Oligonucleotides SRE and SREm1 (which does not include the E-box) as well as procedures for EMSA analysis were as previously described¹. The probe was an oligonucleotide containing the MCK 'down stream' canonical E-box (5'-CTAGACCCCAACACCTGCTGCCTT-3' and complementary strand).

include the E-box. Hence a discrepancy with our results is hardly a surprise.

A. Harel-Bellan

Laboratoire de Biologie des Tumeurs
Humaines,
URA 1156 CNRS,
39 Rue Camille Desmoulins,
94805 Villejuif Cedex, France

Black holes in globular clusters

SIR — It has been suggested^{1,2} that the presence of numerous primordial black holes in globular clusters might lead to the formation of close black hole/normal star binaries through tidal capture interactions. A few of these systems might be observable via the detection of transient emission in the hard X-ray energy range ($E > 50 \text{ keV}$), similar to that observed in the accreting black hole binaries outside globular clusters³.

We would like to point out that observational evidence for hard X-ray emission from a globular cluster already exists. A hard X-ray source (35–200 keV) probably associated with the globular cluster Terzan 2 (ref. 4) was discovered using the SIGMA coded mask telescope on the GRANAT satellite⁵.

Terzan 2 is a core-collapsed low-density globular cluster containing a luminous soft X-ray source, X1724–308 (ref. 6). Because type I X-ray bursts have also been observed from Terzan 2 (ref. 7), X1724–308 has been considered to be a low-mass X-ray binary where the accreting object is a weakly-magnetized neutron star.

It is not known whether the hard X-ray emission originates in X1724–308.

Terzan 2 could contain two different sources: a black hole binary responsible for the variable hard X-ray emission observed by SIGMA, and a neutron star binary accounting for the X-ray bursts. Analysis of high angular resolution X-ray (ROSAT) and radio (ATCA) observations, now in progress, as well as future monitoring at high energies with GRANAT, will help to assess the nature of the accreting source(s) in the globular cluster Terzan 2.

Didier Barret

Centre d'Etude Spatiale des
Rayonnements,
CESR, 9 Av. du Colonel Roche,
BP 4346-31029 Toulouse,
Cedex, France

Sandro Mereghetti

Istituto di Fisica Cosmica del CNR,
v. Bassini 15, Milano,
20133 Italy

1. Trouche, D. *et al.* *Nature* **363**, 79–82 (1993).
2. Metz, R. & Ziff, E. *Oncogene* **6**, 2165–2178 (1991).
3. Blackwell, K. T. & Weintraub, H. *Science* **250**, 1104–1109 (1990).
4. Sun, X.-H. & Baltimore, D. *Cell* **64**, 459–470 (1991).
5. Schroter, H., Mueller, C. G. F., Meese, K. & Nordheim, A. *EMBO J.* **9**, 1123–1130 (1990).
6. Walsh, K. *Molec. cell. Biol.* **9**, 2191–2201 (1989).
7. Tuil, D. *et al.* *J. molec. Biol.* **213**, 677–686 (1990).
8. Santoro, I. M. & Walsh, K. *Molec. cell. Biol.* **11**, 6296–6305 (1991).
9. Treisman, R. *Trends biochem. Sci.* **17**, 423–426 (1992).

1. Kulkarni, S. R., Hut, P. & McMillan, S. *Nature* **364**, 421 (1993).
2. Sigurdsson, S. & Hernquist, L. *Nature* **364**, 423 (1993).
3. Tanaka, Y. *Proc. 23rd ESLAB Symp.* p. 3 (eds Hunt, J. & Battrick, B.) ESA SP-296 (Noordwijk, 1989).
4. Barret, D. *et al.* *Astrophys. J.* **379**, L21 (1991).
5. Paul, J. A. *et al.* *Adv. Space Res.* **11**(8), 289 (1991).
6. Grindlay, J. E. *et al.* *Astrophys. J.* **282**, L13 (1984).
7. Grindlay, J. E. *et al.* *Astrophys. J.* **240**, L21 (1980).

Military imbalance

William H. McNeill

A History of Warfare. By John Keegan. *Hutchinson/Knopf*. 1993. Pp. 432. £20/\$27.50.

THIS learned and provocative book is not a history in the ordinary sense. Rather than following a single chronological path, it is structured thematically around what Keegan believes to be three distinct styles of warfare: primitive, Oriental and Western. The book is also a sustained attack on Karl von Clausewitz (1780–1831), the Prussian general whose myopic theories, Keegan believes, contributed to a militarization of European and other societies that threatens human survival.

Keegan seems profoundly ambivalent about his subject: "Warfare is almost as old as man himself, and reaches into the most secret places of the human heart, places where self dissolves rational purpose, where pride reigns, where emotion is paramount, where instinct is king." His very first words are: "I was not fated to be a warrior", owing, he explains, to physical disability. Nonetheless, he feels the "allure that the warrior life exerts over the male imagination". For, as he goes on to explain, "most soldiers are satisfied merely by the company of others, a shared contempt for the softer world, by the liberation from narrow materiality brought by the camp and the line of march, by the rough comforts of the bivouac, by competition in endurance, by the prospect of *le repos de guerrier* among their waiting womenfolk". Yet the way of the warrior is now obsolescent. "There is no precedent . . . for the menace with which future war now confronts the world. Charting the course of human culture through its undoubtedly warlike past towards its potentially peaceful future is the theme of this book."

Keegan accordingly charts human culture through five chapters ("War in History", "Stone", "Flesh", "Iron" and "Fire"), divided from one another by four brief (and often brilliant) 'interludes' in which he takes a global view of the limitations on warmaking, fortification, armies, and logistics and supply. In his first chapter, he argues the case against Clausewitz by resorting to anthropology. In the author's own words: "To look beyond military slavery into the even stranger military cultures of the Polynesians, the Zulus and the samurai, whose forms of warfare defied altogether the rationality of politics as it is understood by Westerners, is to perceive how incomplete, parochial and ultimately misleading is the idea that war is the continuation of politics."

"Stone" deals with primitive and

early historic warfare. What distinguishes primitive war, Keegan believes, is that organized killing is embedded in, and restrained by, ritual. He then — almost a third of the way through the book — takes up the historic beginnings of warfare (first firmly attested by the walls of Jericho, built about 7000 BC) and its manifestations in ancient Sumer and Egypt.

"Flesh" is devoted to the extraordinary

mutuality between horse and warrior".

"Iron" embraces Greek, Roman and mediaeval European warfare. As before, Keegan asks himself what each age contributed to twentieth-century warfare. The Greeks, he argues, placed "the highest premium" on "settling matters as quickly and decisively as possible". Commitment to pitched battle is therefore what modern Europeans learned from Marathon and Plataea. The Romans, for their part, contributed the idea and ideals of the military as a profession. Mediaeval Europe added the idea of holy war and monkish military orders in whose practices "we may perceive the origins of the regimented armies that arose in Europe in the sixteenth century".

"Fire" covers gunpowder and the technological revolution its introduction into

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Japanese warriors. The cult of the sword in Japan held the gunpowder revolution at bay until the nineteenth century.

role of horses in Eurasian warfare. The story begins with the invention and diffusion of war chariots, beginning about 1700 BC, and took a new form with the cavalry revolution after 700 BC. Keegan then sketches the historic career of steppe warriors from Attila to Genghis Khan, aiming always to distil the principles of their style of warfare which, Keegan thinks, remained the same across more than a millennium. What really concerns him, however, is how encounters with steppe peoples influenced European warfare, as elements derived from the steppe contributed to making the Western way of war uniquely formidable and destructive in modern times. These elements included "long-range campaigning, high-speed manoeuvre on the battlefield, effective missile technology, the application of the wheel to warfare and, above all,

European warfare precipitated. Keegan is loath to admit that the Chinese pioneered gunpowder weapons and even suggests that Europeans may have invented gunpowder independently. Such views are now generally discredited, thanks to Joseph Needham's magisterial researches. But Keegan appears to be ignorant of Chinese military history, treating it as an adjunct to steppe warfare and part of what he calls "Oriental warmaking", characterized by "evasion, delay and indirectness". This is a completely inadequate description of a military tradition that was comparable in complexity and success to anything Europe knew before about 1700.

As for Europe, where he is truly at home, Keegan paints the course of warfare in the gunpowder age in dark colours. For instance, he characterizes the professional soldiers of Old Regime armies as

victims of "a military slave system" and sees the wars of the French Revolution as militarizing society "from top to bottom". This allows him to resume his quarrel with Clausewitz and to deplore the fashion in which "Revolutionary weapons, the warrior ethos and Clausewitzian philosophy . . . ensured that, under Hitler's hand, warmaking in Europe between 1939 and 1945 achieved a level of totality of which no previous leader — not Alexander, not Genghis, not Napoleon — had ever dreamed". That, he argues, is what we must now repress, and he concludes his book with the words: "The habits of the primitive — devotees themselves of restraint, diplomacy and negotiation — deserve relearning. Unless we unlearn the habits we have taught ourselves, we shall not survive."

Keegan, in short, wants to use history and anthropology to cure the ills of our time. It is an admirable ambition, but involves a selective sensitivity to the past that seriously distorts things as they were. In particular, his schematization of war as primitive, Oriental and Western strikes me as inadequate: Chinese, Indian and

Islamic warfare cannot be lumped with steppe warfare into a single, undifferentiated whole. I suspect, too, that what Keegan calls primitive war differed so much from case to case that the commonality of ritual scarcely matters. After all, ritual still envelops the life of British regiments in 1993. Does that make them primitive?

But conceptual inadequacy does not diminish readability. Keegan presents a lot of fascinating information, and salts his pages with pithy sayings. Above all, the book is informed by a high and serious concern about the current dilemmas of our Western way of war. The subject is important; Keegan has thought about it long and hard; and his command of European (especially modern European) military history is masterly. He deserves respectful attention, even if his effort to embrace the whole wide world falls severely short. □

William H. McNeill, emeritus professor of history in the University of Chicago, is at 36 School House, Colebrook, Connecticut 06021, USA.

Environmental well-being

David W. Pearce

The New Protectionism: Protecting the Future Against Free Trade. By Tim Lang and Colin Hines. *Earthscan*: 1993. Pp. 184. £10.99 (pbk).

THE theory of free trade asserts that the global community is generally better off if it is allowed to exchange goods and services without protectionist hindrances such as tariffs and quotas, or artificial incentives such as subsidies. Protection does enable any one country to gain, initially anyway, at the expense of others. The essence of the protection versus free trade argument, then, is how to devise incentives for countries to cooperate to maximize global well-being. This requires a set of rules and regulations aimed at mutual assurance against protectionism such as those being developed, albeit slowly, under the various negotiating rounds of the General Agreement on Tariffs and Trade. Now free trade is under attack from a new direction. Some environmentalists argue that free trade damages the environment. Free trade must therefore end in the name of environmental quality. Tim Lang and Colin Hines's book is one example of this environmental critique of free trade.

Unfettered free trade does have its problems. It says nothing about the fairness or otherwise of a Sri Lankan tea-worker's wages, for example. Because it is based on actual costs of production, free

trade need not be fair trade. Moreover, unless there are environmental regulations in place, free trade ignores the environmental impact of production and transport. This impact can be minimized through regulations or, better still, the use of market-based economic instruments such as environmental taxes or tradeable resource quotas. Lang and Hines rightly call for such measures. This is not new of course. It has been a feature of environmental economics for at least three decades. Had the authors rested their case at this point they might have retained some credibility. Unfortunately, they go further.

The authors argue that free trade itself must end or, at least, must be confined to what is 'necessary' only after communities meet their own needs, or trade with other local communities. This is the 'new protectionism', a thin disguise for self-sufficiency. It is never made clear what the natural self-sufficient unit is. Because international trade is an extension of personal trade, perhaps each individual should be self-sufficient, trading with others only when 'necessary'. Presumably some benign dictator determines what is a necessary infringement of the freedom to trade.

The authors are confused on many issues. Only a few examples can be given. They say that free trade encourages economic growth and that this harms the environment. But this confuses growth

of incomes with growth of materials and energy use. The two are linked, but the ratio between them is subject to substantial manipulation. The authors deny this possibility in one section of the book, but embrace it when attempting to refute the argument that localized, self-sufficient trade entails high costs of production due to the loss of economies of large-scale markets. Their argument becomes old-fashioned anti-growthism. The implications for poverty and unemployment are ignored. The authors' protest that their protectionism is not 'classical' protectionism is hardly persuasive.

The focus on trade as a 'cause' of environmental decay also severely distorts the true perspective. There are many factors at work, including overvalued currencies, massive subsidies to agriculture and energy, excessive population growth and insecure or nonexistent property rights. The authors rail against big business, the banks, United Nations organizations and capitalism generally, sometimes to good effect. This may be politically correct environmentalism, but it is hardly balanced research. One looks in vain for criticism of corrupt and misguided government interventions.

The book is replete with mistakes, mainly because the authors are ill at ease with economics. The conflict between wildlife and cattle in Botswana, for example, is not an example of the evils of free trade, but of protection. Botswana's beef industry is encouraged by preferential European Community tariffs. Indonesia's ban on exporting logs, of which the authors approve, had nothing to do with environmental protection but everything to do with capturing downstream value added from wood processing. Unfortunately, because this industry is inefficient, the ratio of trees to timber product is high. More, not less, deforestation has been encouraged.

In the final chapter, the authors attempt to head off expected criticism, but only make matters worse. They state that free trade has actually worsened human well-being, but there is not a single reference to the literature that has tried to estimate the gains from trade. Indeed, the book is remarkable for ignoring most of the rigorous research on the trade and environment issue. That literature shows that there is an area for legitimate debate. This volume should have been an opportunity to contribute to it. As it stands, however, it is a gift to the narrow advocates of free trade for, if this is the best that environmentalists can do, free trade rules OK. That would be a shame. □

David W. Pearce is in the Centre for Social and Economic Research on the Global Environment, University College London, Gower Street, London WC1E 6BT, UK.

The whitewashing of an Englishman

Peter J. Bowler

Sir Francis Galton, FRS: The Legacy of his Ideas. By Milo Keynes. Macmillan: 1993. Pp. 237. £40.

FRANCIS Galton (1822–1911) was Charles Darwin's cousin and a polymath with interests in heredity, evolution and the inheritance of intelligence. He was the founder of the eugenics movement, which aimed to improve the character of the human species by selective breeding. The dust-jacket proclaims this collection, the proceedings of a conference held at the

fact that his ideas on the inheritance of intelligence turned out to have appalling consequences when taken up by those who were determined to suppress the multiplication of what they considered to be inferior characters.

The only piece in the book to offer an evaluation of the substantial historical literature on Galton and the eugenics movement is by W. F. Bynum. Most of the other studies show either no knowledge whatsoever of previous literature on the topic, or at best only dim appreciation of the issues raised by the historians who have already tried to evaluate Galton's influence. To be fair, there has been little discussion of Galton's work in some areas; the papers on his geographical and meteorological interests and his work on fingerprints therefore offer useful surveys. But in other areas the authors' lack of historical sophistication shows up all too obviously. A. W. F. Edwards, writing on statistics, points out quite correctly that Galton's influence on Karl Pearson has been exaggerated. Edwards does at least provide a comprehensive bibliography, but he does not really tackle some of the controversial claims made about the origins of statistics, including Donald Mackenzie's views on the role of social factors in shaping both Galton and Pearson's approaches to the study of large populations.

Other authors do not even get this far. John Maynard Smith, writing on Galton and evolution, does at least admit that he is writing Whig history and points out that, as a scientist, he cannot be expected to approach the job any other way. Fair enough; but why then choose a biologist, however eminent, rather than an experi-

enced historian of biology to write on this important area? An experienced historian would have known, for instance, the background to Galton's belief that evolution occurred through saltations. Why choose a sociologist, Michael Banton, whose concern for modern attitudes towards race prevents him even from facing up to the fact that, by almost any definition, Galton (like most of his contemporaries) was a racist? Those who doubt that Galton thought that African and Australian blacks were different subspecies with lower intellectual and moral standards than whites should read the chapter on "The Comparative Worth of Different Races" in his *Hereditary Genius*. Why choose a psychologist, H. J. Eysenck, who scarcely pauses to pay even lip-service to Galton's memory before rushing into an account of modern ideas on the inheritance of intelligence?

As the examples of Ernst Mayr and Stephen Jay Gould show, scientists can do good history of science when they take the pursuit seriously. But all too often it is assumed that simply being a scientist qualifies one to write authoritatively on the history of one's own field. This book reveals the fallacy of that assumption, and suggests that the assumption is all the more dangerous when the subject matter is potentially controversial. Despite recent cutbacks, there are still plenty of historians of science who could have written serious evaluations of the topics in this volume. It's a pity they were not consulted. □

Peter J. Bowler is in the Department of Social Anthropology, Queen's University of Belfast, Belfast BT7 1NN, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

National Library of Medicine/SPL

Roots of a frontier world

Peter Atkins

The World of Physical Chemistry. By Keith J. Laidler. Oxford University Press: 1993. Pp. 476. £55, \$85.

THE range of a subject is determined by its history. Laidler's story of physical chemistry extends from its beginnings, when numbers were first attached to matter, to the edge of the present when the subject stretches from one coast of chemistry to the other. Physical chemistry is essentially the framework of chemistry, providing the principles of the subject and a vehicle for the interpretation of the properties of matter created in the laboratories of inorganic and organic chemists. It juts out from the bedrock of physics, dealing with the physical properties of atomic and bulk matter, the rates at which reactions change one substance into another, the thermodynamic properties of matter that

determine its equilibrium properties, and the spectroscopic techniques that provide chemists with detailed information about the properties and structures of individual molecules. Even nonspecialist readers will find much to enjoy in the first chapter, which explores the broad sweep of the subject and looks at the impressive contributions that it has made at the interface of pure research and technology. Fox Talbot, for instance, did detailed work on spectroscopy before developing his photographic process.

Laidler, in his carefully researched and scholarly volume, spans the development of physical chemistry from its halting beginnings with Boyle and Henry, when the properties of the simplest phase of matter were first tamed, through the atomists, such as Dalton and Avogadro, who first provided the essential insights on which all current interpretations of the

Galton Institute in London, to be "the definitive book on Francis Galton". Unfortunately it is nothing of the kind, partly because too many of the contributors simply do not have the kind of historical expertise needed to evaluate the work and influence of a major figure such as Galton.

Surprisingly, there is no specific discussion of Galton's ideas on the actual process of inheritance, including his "law of ancestral heredity". More seriously, although there are papers on various aspects of Galton's efforts to come to grips with human heredity, including twin studies, there is no serious discussion of the pernicious effects of the eugenics movement in the early twentieth century. Perhaps this was inevitable given the location of the original conference, but one can hardly call a book on Galton's legacy 'definitive' when it fails to confront

macroscopic in terms of the microscopic are based, to the giants of the late nineteenth and early twentieth centuries, such as Gibbs, van't Hoff, Ostwald and Nernst. The account stops short of today's work, and is perhaps best seen as a survey of roots up to the mid-twentieth century, apart from gestures in the direction of modern kinetics, where chemical rates on a femtosecond timescale are starting to be explored. We see how far the subject has progressed not only from its beginnings in Oxford in the seventeenth century but also since the nineteenth century when spectra were published as woodcuts.

There are discernible themes, not always consciously exploited but a pleasure (and sometimes a concern) to detect. Where, for instance, are the women of physical chemistry? In so far as diffraction is not accepted as a part of physical chemistry (a questionable but understandable omission), then the barrel of the subject must be scraped quite closely before female physical chemists stand up to be counted. Laidler awards 'Box' status (in the sense that he uses one-page boxes for short biographies of outstanding and mostly dead contributors to the field) to only two women. I must confess I had heard of neither of them, but both, and especially Elizabeth Fulhame (who was alive towards the end of the eighteenth century and into the early nineteenth), sound like remarkable women. Of the other, Agnes Pockels, Lord Rayleigh was moved to wonder whether the name Agnes given to a German necessarily implies a woman.

Another theme is the regularity with which scorn greeted many of the great when they first faltered out into the world. Avogadro was doubted, and Arrhenius was awarded the lowest class of doctorate for proposing the dissociation of electrolytes. So too was de Broglie granted a minor class of degree. In a pre-echo of a remark made later in another context, van de Graaff said, when commenting on de Broglie's public examination, that never had so much gone on over the heads of so many.

The other side of this coin is the regrettable failure to acknowledge the greatness of some seminal contributions to the subject with the Nobel prizes they so properly deserved. It is inexcusable that G. N. Lewis was never recognized in this way (perhaps because of Bohr's influence and his distrust of the concept of electron pairs which seemed to be too stationary to fit into the concept he had formed of atoms). It is also inexcusable that Henry Eyring was never recognized. Each will take comfort, in their respective heavens, from the recognition that Laidler's book will provide on their behalf.

Then there are the recognized but modest practitioners, who were honoured

publicly in their time by all the awards that were appropriate, but who still kept some of their light under bushels. Of these, perhaps the most outstanding is van't Hoff, who as well as achieving so much that we teach our students, also proposed equations and concepts that we commonly ascribe to others, such as the Arrhenius equation for the temperature dependence of reaction rates and the principle we commonly attribute to Le Chatelier. Van't Hoff's modest charm comes through on Laidler's pages.

Here too are the eccentrics who are to be found in any subject but who perhaps are unduly numerous in this mongrel field on the frontiers of physics and chemistry. We meet Henry Armstrong and Percival Pickering, whose work has left no memorial except the lesson drawn by the latter's obituarist, who said that he had all the spare parts of genius, but not the patience to put them together.

Laidler has provided a masterly survey of the field, which will help to put the record straight in many backwaters of the subject and in not a few estuaries too. □

Peter Atkins is at Lincoln College, Oxford OX1 3DR, UK.

Helping hands

Russell H. Tuttle

Hands of Primates. Edited by H. Preuschoft and D. J. Chivers. Springer: 1993. Pp. 421. DM198, \$136, £78.50.

ALTHOUGH not vital, our hands are precious organs indeed. They are rivalled only by the brain, speech apparatus and bipedal locomotor complex as foci of research among comparative morphological primatologists and palaeo-anthropologists.

Preuschoft and Chivers are to be commended for assembling this lavishly illustrated collection of 25 papers on hand use, functional morphology and evolution and ontogeny written predominantly by Western European experts. This is not a book for the general reader, and even fellow researchers will be confounded by the absence of an index, which would direct them to particular points of interest. Myriad infelicities of English are surprising considering that the second editor is British.

Nonetheless, I recommend that *Hands of Primates* be added to local libraries and that anybody who studies primate hands should scan its pages for the many gems of information that it contains. I recommend also that those who model primate, and particularly hominid, evolution should read the papers by C. Boesch and co-workers, based on their invaluable studies

of unprovisioned chimpanzees in the Tai Forest, Ivory Coast, and by M. Godinot, C. Beard and co-workers on the fossil record of primate manual morphology.

M. Günther and C. Boesch estimate that Tai chimpanzees gain from coula nuts about nine times the energy that they expend to collect, crack and ingest them. Coula and several tougher species of nuts are not readily exploited without proper tools and high-precision tool behaviour. They are a major resource for Tai chimpanzees during a four-month period each year when other staple foods are scarce.

The remarkable manipulatory skills of pongid apes, based on independent fine control of the fingers, are emphasized by C. Boesch and H. Boesch and M. E. Redshaw. Apparently, the marked morphological compromises, which are related to knuckle-walking, versatile climbing and other positional behaviours, are not insuperable impediments to precise manipulation among chimpanzees, bonobos, gorillas and orangutans. Accordingly, apish features in the hands of early Hominidae — *Australopithecus afarensis*, *Paranthropus* sp., and *Homo habilis* — do not preclude them from using tools. Indeed, as I have long maintained, the gross manual morphologies of extant apes, Old World monkeys and many New World monkeys could accommodate the manufacture and use of Oldowan stone tools and crude vegetal tools.

Interestingly, Tai chimpanzees employ smaller stones arboreally to hammer nuts than they do terrestrially. This fact could be useful to infer possible arboreal tool use by early hominids at Olduvai Gorge, Tanzania, and elsewhere.

Although fossil hand bones are being recovered in increasing numbers and with better contextual control, we remain quite ignorant of the precise manual morphologies that were ancestral to the dazzling array of special and more mundane hands of extant primate species. For instance, Beard and colleagues argue compellingly that the hands of *Proconsul*, an eastern African Miocene dental ape, evince no features of knuckle-walkers or dramatic arm-swingers, but instead indicate generalized arboreal quadrupedalism and climbing. Although the gross manual complex of *Proconsul* is quite suitable as a template on which natural selection could act to produce the hands of knuckle-walkers, versatile climbers and brachiators, the phylogenetic links between *Proconsul* and extant African apes (*Pan* spp.) and orangutans (*Pongo pygmaeus*) are frustratingly penumbral. □

Russell H. Tuttle is in the Department of Anthropology, University of Chicago, 1126 East 59th Street, Chicago, Illinois 60637, USA.

Period-doubling reversals and chaos in simple ecological models

Lewi Stone

The period-doubling route to chaos is a well known feature of a range of simple, nonlinear difference equations routinely used in modelling biological populations. It is not generally understood, however, that the process may easily break down and suddenly reverse, giving rise to distinctive period-halving bifurcations. These reversals may act to control, and possibly prevent, the onset of chaos.

THEORETICAL notions of balance, stability and change have all been instrumental in shaping our perceptions of ecological and biological processes^{1,2}. Even the earliest human societies developed some rudimentary formulation of the background assumption of ecology: the 'balance of nature' concept¹. The belief in a stable, ordered equilibrium world, in which nature is normally considered to exist in balance, generally remained unchallenged for millennia. Yet the concepts of imbalance, variability and indeterminate chance events proved to be of crucial importance, so much so that by 1930 the celebrated British ecologist Charles Elton could safely state that: "'The balance of nature' does not exist, and perhaps never existed. The numbers of wild animals are constantly varying to a greater or less extent, and the variations are usually irregular in period and always irregular in amplitude"³. It took decades for the ecological community to digest Elton's message, and even as late as 1980, Simberloff found it necessary to stress the significance of probabilism, arguing that what might be noise to the physicist "is music to the ecologist"⁴.

It was not until 1974, however, that Robert May demonstrated that this seemingly stochastic behaviour could in fact be explained by the highly complex and chaotic behaviour that arises from nonlinearities in the simplest deterministic population models, an insight which has had a radical impact both in ecology and elsewhere.

Here I review some of the basic concepts pertaining to chaos in population models and reexamine the conditions under which it occurs. The period-doubling route to chaos (described below), thought to be generic to the family of 'single-humped' one-dimensional maps, may in important cases be absent. For many models, a seemingly minor structural perturbation with real ecological justification is all that is required to break down and reverse the expected period-doubling route to chaos. If these models are taken to illustrate population dynamics in the real world, then chaos may be a fragile process, the well known 'universal' behaviour being easily inhibited. More importantly, these period-doubling reversals may be used to control chaos, as they have the potential to suppress dangerous chaotic fluctuations.

First-order equations

Simple nonlinear difference equations provide a unifying framework for a diverse array of mathematical models encompassing a wide range of disciplines⁵⁻¹⁰. Without loss of generality, the following study is set in an ecological context where the variable N_t represents the population size or biomass of a species in generation t , and satisfies a general first-order equation of the form:

$$N_{t+1} = F(N_t). \quad (1)$$

Here F is a one-dimensional map reflecting nonlinear density-dependent growth⁵. Unless otherwise stated, it is assumed that F is 'single-humped' (possessing a unique maximum) with a shape

controlled only by one parameter, such as r . The time-evolution of the population model is found by iterating the equation from an initial population $N_1 > 0$, to yield the sequence:

$$N_1 \rightarrow N_2 = F(N_1) \rightarrow N_3 = F(N_2) \rightarrow N_4 = F(N_3) \rightarrow \dots$$

Under these bare specifications, equation (1) displays an intricate structure of cyclical solutions governed only by the parameter r (Fig. 1).

When F belongs to a general class of 'single-humped' functions, termed here U -functions, equation (1) displays a bifurcation structure that is 'universal' in that: first, the bifurcation structure conforms with Feigenbaum's period-doubling route to chaos whereby the number of stable periodic points successively double as the parameter r increases, in a manner that is ultimately independent of the specific form of the U -function⁹; second, as r approaches a critical value r_c , an infinite number of fixed points emerge and the population dynamics become chaotic; and third, when r is increased beyond r_c , a chaotic regime is entered that contains an ordered U (universal) sequence of periodic windows¹¹.

Reversals in a model of immigration

Surprisingly, the precise conditions that determine whether F actually belongs to the special class of U -functions, and so inherits these properties, are unknown^{11,13}. But it has been suggested that to exclude all cases of anomalous non-universal behaviour would necessarily "restrict the underlying class of transformations [U -functions] rather drastically"¹¹. There is thus a need to assess whether U -functions are general enough to represent ecological processes usefully, and to question how robust their dynamics might be to small but realistic structural changes.

To gain some understanding of the problem, consider the well known model of logistic population growth:

$$N_{t+1} = F(N_t) = N_t \exp [r(K - N_t)] \quad (2)$$

which displays all the features characteristic of a U -function. Equation (2) has become a model for illustrating the transition to chaos in one-dimensional maps^{5,7}. In biological terms, the parameter r reflects the population's natural growth rate, and K reflects the carrying capacity of the environment and may be scaled to unity ($K=1$) without loss of generality.

This model can be extended to allow for other ecological factors. Numerous sub-populations relying either on refuges for protection, or on an influx of individuals by immigration, must for example have a floor ($\lambda > 0$) below which the population level never falls. To model this, McCallum⁸ modified equation (2) as follows:

$$N_{t+1} = N_t \exp [r(1 - N_t)] + \lambda, \quad \lambda > 0 \quad (3)$$

with $N_t > \lambda$. The constant term λ is intended to represent either the number of immigrants per generation or, if a refuge exists, that small sub-population which becomes isolated and therefore largely released from density-dependent effects.

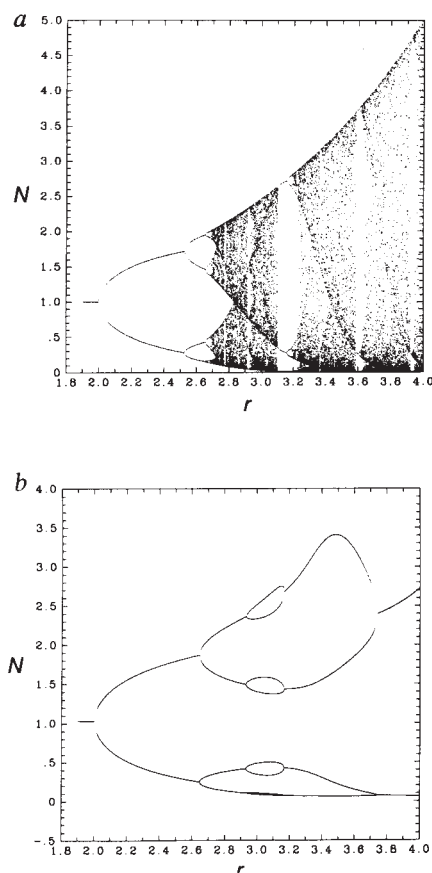


FIG. 1 Bifurcation diagram for *a*, equation (2); *b*, equation (3), $\lambda = 0.06$. For each of 1,100 values of *r*, the respective equation was iterated 700 times and the last 200 points were plotted. The bifurcation diagram in *a* summarizes the periodic and aperiodic dynamics of the logistic population model:

$$N_{t+1} = F(N_t) = N_t \exp [r(K - N_t)]. \quad (B1)$$

To review briefly some of the principal ideas involved, first note that the iterates of the model form a *p*-cycle of period *p*, if for some *t*, $N_t = N_{t+p}$ but $N_t \neq N_{t+j}$ for any integer *j* where $1 \leq j < p$. The model then repeatedly cycles through *p* distinct points and the stability of the cycle can be deduced from standard analytic or numerical techniques^{6,7}. For values of $r < 2$, one sees from the bifurcation diagram that the equation possesses an attracting equilibrium point. By increasing *r* to a value slightly greater than $r = 2$, the equilibrium bifurcates and a two-cycle is born; the model then alternates between two population levels $N_1 \rightarrow N_2 \rightarrow N_1 \rightarrow N_2 \rightarrow \dots$. By raising *r* yet further, the two-cycle bifurcates and forms a four-cycle $N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_4 \rightarrow N_1 \rightarrow N_2 \rightarrow \dots$. Increasing *r* in this same manner, yields an infinite succession of bifurcations so creating a hierarchy of stable cycles of periods 2, 4, 8, 16, ..., 2^n ($n \rightarrow \infty$). This succession of bifurcations develops at a geometric rate and accumulates asymptotically as $n \rightarrow \infty$ at $r = r_c$. For the above logistic map $r_c = 2.6924$ where aperiodic dynamics begin. Beyond this point of accumulation the bifurcation diagram displays a highly complicated chaotic regime interspersed with periodic windows. Each window is characterized by its own specific periodicity (that accords with the *U*-sequence¹¹) and period-doubling cascade. For example, a period-3 window is clearly visible in the bifurcation diagram, beginning at $r = 3.1024$. An enlargement of the window would reveal a period-doubling cascade with a hierarchy of cycles of period $3, 3 \cdot 2^1, 3 \cdot 2^2, 3 \cdot 2^3, \dots, 3 \cdot 2^n$ ($n \rightarrow \infty$).

For values of $\lambda \ll 1$, equation (3) might be viewed as a structural perturbation of equation (2). But the two models possess very different and qualitatively distinct dynamics. Figure 1*b* displays the bifurcation diagram of equation (3) when $\lambda = 0.06$. It explicitly illustrates a form of 'bubbling'¹³ in which the expected period-doubling bifurcation process curtails and reverses, giving

rise to period-halving bifurcations. Successive orbits of periods 1, 2, 4, 8, 4, 2 occur as *r* increases from zero. Increasing small period nonlinearity (*r*), and thus the severity of density-dependence, ultimately tends to stabilize the oscillations rather than producing chaotic population dynamics. Numerical calculations find no indication of stable cycles with a period greater than eight. On comparing the diagrams in Fig. 1, it is clear that this phenomenon of period-halving is not observed in equation (2). Furthermore, the anomaly of period-doubling reversals or bubbling derives from structurally perturbing what has been understood to be a *U*-function, in this case by an amount that is only 6% of the total carrying capacity.

A more complete investigation reveals the extent to which λ suppresses the period-doubling route to chaos. Figure 2 outlines the behaviour of equation (3) over a range of values in *r*- λ parameter space and demarcates regions in which the model exhibits stable orbits of periods $p = 1, 2, 4$ or > 4 . The model population is generally found either at equilibrium ($p = 1$) or characterized by a stable 2-cycle. Only a relatively small region of parameter space represents chaotic behaviour.

General occurrence of period-doubling reversals

Analysis of a number of common difference equation models reveals that perturbations of the type used above often lead to the suppression of chaos and the initiation of a reversal with period-doubling, eventually followed by period-halving as the parameter *r* is increased. The phenomenon occurs in models of the crown-of-thorns starfish⁸, insect populations^{14,15}, perennial grass inhibited by plant litter¹⁶, annual plant populations¹⁷, host-parasitoid interactions¹⁸, and systems of competing species⁷. The effect itself has already been observed in several models of genetic selection¹⁹, microbial predator-prey chemostats²⁰, phytoplankton-herbivore interactions (J. Steele and M. Pascual, personal communication), cardiac cell stimulation²¹, and coevolutionary host-parasite models²², but has generally not been properly interpreted and has almost never been framed in terms of the single-hump one-dimensional map (but see ref. 23). The presence of reversals has also been documented in other areas of research, ranging from models of magnetoconvection²⁴ and rotating galaxies²⁵, to a neuronal model of psychotic human behaviour²⁶. Also notable is Swinney's classic study^{27,28} of a chemical reaction in a stirred flow reactor yielding important empirical confirmation.

The explanation for this reversal of period-doubling in a broad class of function *F* is simple. As demonstrated in Box 1, the phenomenon may arise whenever a single-humped map has a point about which the function flattens out into a 'plateau'. Bifurcations are governed first by the 'hump', which is largely responsible for period-doubling; and second, by the concavity of the 'plateau' and its height (above the *x*-axis), factors which are conducive to period-halving. In some cases, the same argument can be extended to the analysis of one-dimensional maps

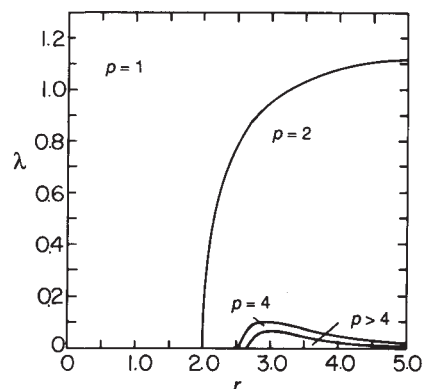


FIG. 2 Regions in *r*- λ parameter space for equation (3) characterized by low-order stable cycles of periods $p = 1, 2, 4, > 4$ as determined numerically.

BOX 1 The reversal of a period-doubling cascade

To explain how a period-doubling cascade reverses, it is useful to examine the graph of the single humped function F which has been scaled by various values of the parameter r for analysis of the equation $N_{t+1} = rF(N_t)$. F has a plateau region that flattens out for large N_t . The plateau can occur whenever there is a point of inflection or if the second derivative $F''(N)$ suddenly changes sign to the right of the critical point. The fixed points N^* may be found graphically at the intersections of the 45° line $N_{t+1} = N_t$ and the curve $rF(N_t)$. The stability of N^* is ensured if the slope $S(N) = rF'(N)$ of the curve at the intersection point satisfies $|S(N^*)| < 1$ (refs 5, 6).

Suppose the parameter r is continuously increased from zero. When the nontrivial stable fixed point $N = N^* > 0$ appears (at which point $S(N^*) > 0$), increasing r further serves to reduce $S(N^*)$, until $S(N^*) = -1$, when a period-doubling bifurcation occurs⁴⁰; the fixed point N^* becomes unstable and a stable two-cycle is born. A succession of period-doublings may follow in the usual manner as r is gradually increased. Consider now the case when r is large. Graphically the equilibrium point $N = N^*$ is stable because it lies in the plateau region where $S(N^*) \approx 0$. (The existence of this single stable equilibrium depends chiefly on the concavity and height of the plateau in a manner which can be expressed rigorously. For example, equation (3) has a plateau that is too steep to permit a stable equilibrium, even as $r \rightarrow \infty$, where it instead exhibits a stable two-cycle.) As r is gradually reduced, $S(N^*)$ continuously decreases, until $S(N^*) = -1$, when a period-doubling bifurcation occurs; again the fixed point N^* becomes unstable and a stable two-cycle emerges. As r is reduced further a succession of period-doubling can similarly result.

In summary, as r increases from zero the phenomenon of period-doubling must eventually be followed (at least for large r) by period-halving. (The breakdown in period-doubling occurs when for some stable k -cycle $\{N_i\}_{i=1 \dots k}$, the slope of the k -fold map $F^k(N_i)$ is constrained so that $S(N_i) = F^k F'(N_i) > -1$ and a period-doubling bifurcation cannot take place.) For similar reasons, perturbations that serve to translate the plateau vertically upwards yield the same qualitative result.

The function F in this example has been chosen as 'spiked' only to aid in the visualization of the process whereby period-doubling reversals are initiated. Its spiked shape is not essential to the general argument because vertical perturbations may also considerably enhance or initiate a reversal (as seen in Box 2).

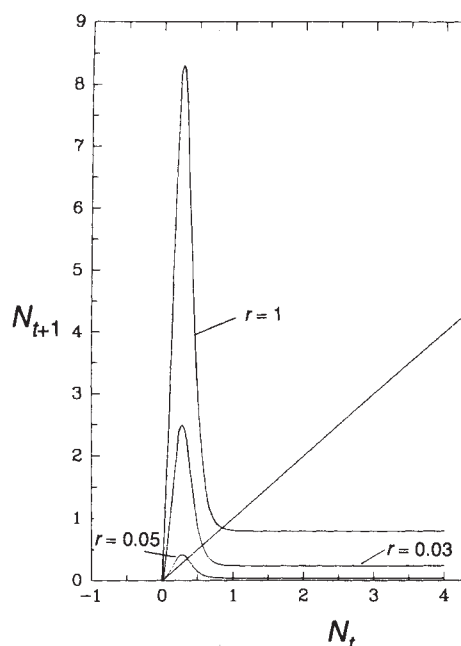
It is instructive to examine the alternative, though perhaps less biologically relevant⁶, quadratic model of logistic growth:

$$N_{t+1} = F(N_t) = rN_t(1 - N_t) \quad (C1)$$

F being a well known U -function. That F is purely concave-down explains in part why it fails to exhibit reversals after being perturbed to the form:

$$N_{t+1} = RN_t(1 - N_t) + \lambda, \quad \lambda > 0 \quad (C2)$$

To see this more directly, note that the perturbed map is, up to a linear change of coordinates, identical to equation (C1), the two maps being connected by the relation $r^2 = R^2 + 4R\lambda$. Hence equation (C2) must also be a U -function and, like equation (C1), follows the same period-doubling route to chaos as R and/or λ increase from zero, without ever exhibiting the phenomenon of period-halving. □

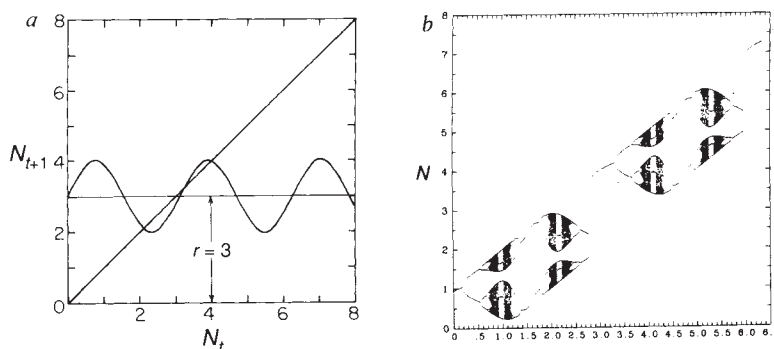


BOX 2 Cardiac chaos

SIMPLE difference equations are also valuable biological models. The complex rhythms of the human heart^{21,29} and the control of chaotic cardiac arrhythmia⁴¹ have been the subject of intense investigation with these modelling techniques. In fact, the two-parameter family of circle maps has been used extensively to simulate cardiac dynamics²⁹. Period-doubling reversals are also a fundamental phenomenon in these models. Consider the degree-zero circle map:

$$N_{t+1} = \sin(\alpha N_t) + r \quad (D1)$$

(This is a slightly simplified form of the usual circle map, but no generality is lost in this context.) Most analyses (but see ref. 42) use numerical techniques to compute the equation's 'skeleton', that is, the locus in parameter space of superstable cycles, those periodic cycles whose trajectory contains the critical point of the map. Instead, here we begin by examining more closely the relationship N_{t+1} versus N_t (Fig. a, with parameters $\alpha = 2$, $r = 3$). Notice that fixed points occur at the intersection of this graph with the 45° line $N_{t+1} = N_t$. For a fixed value of α , increasing r serves to shift the points of bifurcation, weaving them successively through a series of humps and valleys. This can result in a corresponding series of period-doubling reversals with a periodically recurring bubbling structure as seen in the bifurcation diagram of the circle map (with $\alpha = 2$) displayed in b. There is a striking correspondence between this bifurcation diagram and that determined from an empirical model of periodically stimulated embryonic chick heart cells as found in ref. 21. □



with multiple critical points such as the bimodal maps used for modelling insect populations (see below), or the degree-zero circle map used for modelling periodically stimulated biological cells²⁹.

Single-humped functions F that gently flatten out, often exponentially or simply asymptotically, occur in a number of one-dimensional maps supposed to model real systems^{9,14-17,27,28} which are unlikely generally to be U -functions. The class of functions that will not exhibit the 'universal' behaviour described above may thus be large. Furthermore, slight perturbations of a map can augment the effects of a plateau (should one exist) and cause breakdown or distortion of universal behaviour by initiating a period-doubling reversal.

Two-dimensional maps

The reversal of period-doubling cascades in two-dimensional maps surprisingly yields to mathematical analysis far more readily than the one-dimensional case treated above³⁰⁻³². In many of the former maps, reversals are now understood to be inevitable and known to occur infinitely often. 'Antimonotonicity', as the phenomenon has been termed³⁰⁻³², has definite implications for ecological models where higher dimensional maps are often appropriate^{33,34}.

Consider, for example, a simple two-dimensional ecological model (with two time lags):

$$N_{t+1} = RN_t(1 - N_t) - kN_{t-1} \quad (5)$$

where next generation's population is dependent not only on the current population, but also on the previous generation. Viewed in this way, the parameter k might be interpreted as a simple structural perturbation of the basic model.

In fact (as a simple change of coordinates immediately shows), equation (5) is equivalent to the widely studied Henon map, now known to exhibit dramatic antimonotone behaviour; for relatively large parameter regions an infinite number of period-doubling reversals occur³⁰⁻³². Yet, as noted in Box 1, when $k = 0$, equation (5) fails to exhibit period-doubling reversals for any parameter setting. It appears that period-doubling reversals again arise from the presence of a structural perturbation. Further, if this specific example is an indicator, the phenomenon may be more prominent in models of higher dimension.

Dawson *et al.*^{31,32} have argued that one-dimensional cubic maps with two critical points serve as models for antimonotonic-

ity in two-dimensional maps. The argument outlined in Box 1 explains how multiple critical points can initiate period-doubling reversals. The theoretical underpinnings of antimonotone behaviour are extremely complicated and not yet fully elaborated. It is currently a subject of intense research interest.

Discussion

Although physicists have observed several examples of period-doubling in the laboratory, the same cannot be said for ecologists in their studies of communities, where evidence for chaotic dynamics is hard to come by and difficult to confirm³⁵. Theoretical speculations using simple mathematical models are useful in filling this void but, as shown here, do not yield robust conclusions about the incidence of chaos and tend to suggest that it is somewhat fragile and easily inhibited. Moreover, as a recent study has shown³⁶, chaotic dynamics appear to be relatively uncommon in a wide range of simple nonlinear systems for both low-order polynomial maps and ordinary differential equations, when analysed over realistic operational parameter regimes. Taken together, these findings might suggest that the fluctuations and variability we witness in the natural world are probably of stochastic origin, but this conclusion may be premature. The parameters giving rise to complex dynamics may be of little consequence, as it is unclear how readily they can be translated into real-world systems, and the actual ecological validity of some of the better known equations has recently been under dispute^{34,37}. Furthermore, it is unrealistic to expect simple models to correspond, detail for detail, to the intricacies of the very complex natural world^{38,39}. The basic problem rather serves to highlight further the need for theoretical ecologists to better assess the prevalence of low-dimensional chaos.

Of more practical importance, the results reported here indicate that ecological factors such as immigration and predation may significantly control and impede the onset of chaos, thereby suppressing erratic, and possibly dangerous, population fluctuations. The extent to which the phenomenon may occur in real ecosystems awaits future clarification, although McCallum's pioneering study constitutes an important first step⁸. Our ability to gain further insights concerning nonlinearities and chaos in the natural world will depend on further exploration and study of these concepts. □

Lewi Stone is at the Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Tel Aviv, Israel.

- Egerton, F. N. Q. *Rev. Biol.* **48**, 322-350 (1973).
- Wiman, I. M. B. *Ambio* **19**, 62-69 (1990).
- Elton, C. *Animal Ecology and Evolution* (Oxford University Press, New York, 1930).
- Simberloff, D. *Synthese* **43**, 3-39 (1980).
- May, R. M. *Nature* **261**, 459-467 (1976).
- May, R. M. & Oster, G. F. *Am. Nat.* **97**, 573-599 (1976).
- May, R. M. *Science* **186**, 645-647 (1974).
- McCallum, H. I. *J. theor. Biol.* **154**, 277-284 (1992).
- Feigenbaum, M. J. *J. Stat. Phys.* **19**, 25-52 (1977).
- Collet, P. & Eckmann, J. P. *Iterated Maps on the Interval as Dynamical System* (Progress in Physics, Birkhauser, Boston, 1980).
- Metropolis, N., Stein, M. L. & Stein, P. R. *J. combin. Theor. A* **15**, 25-44 (1973).
- Beyer, W. A., Mauldin, R. D. & Stein, P. R. *J. math. Anal. Appl.* **115**, 305-362 (1986).
- Bier, M. & Bountis, C. *Phys. Lett. A* **104**, 239-244 (1984).
- Bellows, T. S. Jr *J. Anim. Ecol.* **50**, 139-156 (1981).
- Hassell, M. P., Lawton, J. H. & May, R. M. *J. anim. Ecol.* **45**, 471-486 (1976).
- Tilman, D. & Wedin, D. *Nature* **353**, 653-655 (1992).
- Palcala, S. W. & Silander, J. A. Jr *Am. Nat.* **125**, 385-411 (1985).
- May, R. M. *Am. Nat.* **125**, 573-584 (1985).
- Altenberg, L. *Am. Nat.* **138**, 51 (1991).
- Kot, M., Saylor, G. S. & Schultz, T. W. *Bull. math. Biol.* **54**, 619-648 (1992).
- Guevara, M. R., Glass, L. & Shrier, A. *Science* **214**, 1350-1353 (1981).
- May, R. M. & Anderson, R. M. *Proc. R. Soc. B* **219**, 281-313 (1983).
- Nusse, H. A. & Yorke, J. A. *Phys. Lett. A* **127**, 328-334 (1988).
- Knobloch, E. & Weiss, N. O. *Physica* **90**, 379-407 (1983).

- Contopoulos, G. *Lett. Nuova Cimento* **37**, 149-153 (1983).
- King, R., Barchas, J. D. & Huberman, B. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1244-1247 (1984).
- Swinney, H. L. *Physica* **70**, 3-15 (1983).
- Coffman, K., McCormick, W. D. & Swinney, H. L. *Phys. Rev. Lett.* **56**, 999-1002 (1986).
- Belair, J. & Glass, L. *Phys. Lett. A* **127**, 113-116 (1983).
- Kan, I. & Yorke, J. *Bull. Am. Math. Soc.* **23**, 469-476 (1990).
- Dawson, S. P., Grebogi, C., Yorke, J. A., Kan, I. & Kocak, H. *Phys. Lett. A* **162**, 249-254 (1992).
- Dawson, S. P. & Grebogi, C. *Chaos Solitons Fract.* **1**, 137-144 (1991).
- Turchin, P. & Taylor, A. D. *Ecology* **73**, 289-305 (1992).
- Berryman, A. A. *Trends Ecol. Evol.* **7**, 316 (1992).
- Stone, L. *Proc. R. Soc. B* **250**, 77-81 (1992).
- Sprott, J. C. *Phys. Lett. A* **173**, 21-24 (1993).
- Ginzburg, L. R. *Trends Ecol. Evol.* **7**, 133 (1992).
- Taylor, P. J. *Oikos* **54**, 121-126 (1989).
- Taylor, P. J. *Oikos* **55**, 434-436 (1989).
- Devaney, R. L. *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, California, 1989).
- Garfinkel, A., Spano, M. L., Ditto, W. L. & Weiss, J. N. *Science* **257**, 1230-1235 (1992).
- Mackay, R. S. & Tresser, C. *Physica* **27D**, 412-422 (1987).

ACKNOWLEDGEMENTS. I thank H. MacCallum, C. Grebogi, H. Nusse, U. Pollinger, M. Kot, E. Tziperman, G. Goodfriend, J.-P. Eckmann, Z. Agur and A. Elitzur for their kind help.

Possible gravitational microlensing of a star in the Large Magellanic Cloud

C. Alcock^{*†}, C. W. Akerlof^{†¶}, R. A. Allsman^{*},
T. S. Axelrod^{*}, D. P. Bennett^{*†}, S. Chan[‡],
K. H. Cook^{*†}, K. C. Freeman[‡], K. Griest^{†||},
S. L. Marshall^{†§}, H.-S. Park^{*}, S. Perlmutter[†],
B. A. Peterson[‡], M. R. Pratt^{†§}, P. J. Quinn[‡],
A. W. Rodgers[‡], C. W. Stubbs^{†§}
& W. Sutherland[†]

^{*} Lawrence Livermore National Laboratory, Livermore, California 94550, USA

[†] Center for Particle Astrophysics, University of California, Berkeley, California 94720, USA

[‡] Mt Stromlo and Siding Spring Observatories, Australian National University, Weston, ACT 2611, Australia

[§] Department of Physics, University of California, Santa Barbara, California 93106, USA

^{||} Department of Physics, University of California, San Diego, California 92039, USA

[¶] Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA

THERE is now abundant evidence for the presence of large quantities of unseen matter surrounding normal galaxies, including our own^{1,2}. The nature of this 'dark matter' is unknown, except that it cannot be made of normal stars, dust or gas, as they would be easily detected. Exotic particles such as axions, massive neutrinos or other weakly interacting massive particles (collectively known as WIMPs) have been proposed^{3,4}, but have yet to be detected. A less exotic alternative is normal matter in the form of bodies with masses ranging from that of a large planet to a few solar masses. Such objects, known collectively as massive compact halo objects⁵ (MACHOs), might be brown dwarfs or 'jupiters' (bodies too small to produce their own energy by fusion), neutron stars, old white dwarfs or black holes. Paczynski⁶ suggested that MACHOs might act as gravitational microlenses, temporarily amplifying the apparent brightness of background stars in nearby galaxies. We are conducting a microlensing experiment to determine whether the dark matter halo of our Galaxy is made up of MACHOs. Here we report a candidate for such a microlensing event, detected by monitoring the light curves of 1.8 million stars in the Large Magellanic Cloud for one year. The light curve shows no variation for most of the year of data taking, and an upward excursion lasting over 1 month, with a maximum increase of ~ 2 mag. The most probable lens mass, inferred from the duration of the candidate lensing event, is ~ 0.1 solar mass.

The MACHO Project^{7,8} uses the gravitational microlens signature to search for evidence of MACHOs in the Galactic halo, which is thought to be at least three times as massive as the visible disk². (Two other groups are attempting a similar search^{9,10}.) If most of our Galaxy's dark matter resides in MACHOs, the 'optical depth' for microlensing towards the Large Magellanic Cloud (LMC) is about 5×10^{-7} (independent of the mass function of MACHOs), so that at any given time about one star in two million will be microlensed with an amplification factor $A > 1.34$ (ref. 5). Our survey takes advantage of the transverse motion of MACHOs relative to the line-of-sight from the observer to a background star. This motion causes a transient, time-symmetric and achromatic brightening that is quite unlike any known variable star phenomena, with a characteristic timescale $t = 2r_E/v_\perp$ where r_E is the Einstein ring radius and $v_\perp$ is the MACHO velocity transverse to the line-of-sight. For typical halo models the time $t \sim 100\sqrt{M_{\text{macho}}/M_\odot}$ days⁵ (where $M_\odot$ is the mass of the Sun). The amplification can be

large, but these events are extremely rare; for this reason our survey was designed to follow >10 million stars over several years.

The survey employs a dedicated 1.27-m telescope at Mount Stromlo. A field-of-view of 0.5 square degrees is achieved by operating at the prime focus. The optics include a dichroic beam-splitter which allows simultaneous imaging in a 'red' beam (6,300–7,600 Å) and a 'blue' beam (4,500–6,300 Å). Two large charge-coupled device (CCD) cameras¹¹ are employed at the two foci; each contain a 2×2 mosaic of $2,048 \times 2,048$ pixel Loral CCD imagers. The 15- μm pixel size corresponds to 0.63 arcsec on the sky. The images are read out through a 16-channel system, and written into dual ported memory in the data acquisition computer. Our primary target stars are in the LMC. We also monitor stars in the Galactic bulge and the Small Magellanic Cloud. As of 15 September 1993, over 12,000 images have been taken with the system.

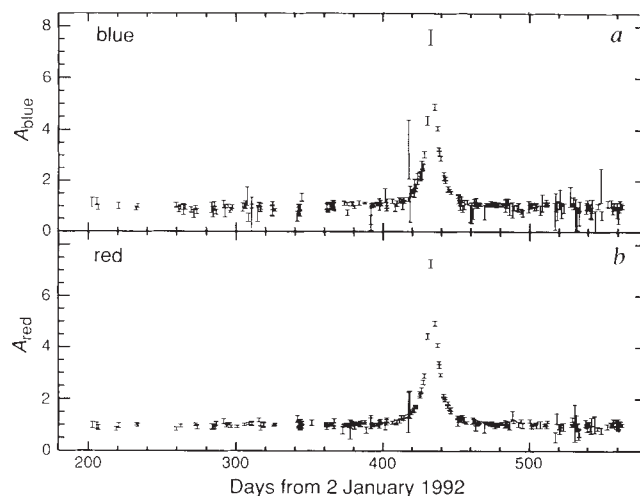
The data are reduced with a crowded-field photometry routine known as Sodophot, derived from Dophot¹². First, one image of each field that was obtained in good seeing is reduced in a manner similar to Dophot to produce a 'template' catalogue of star positions and magnitudes. Normally, bright stars are matched with the template and used to determine an analytic point spread function (PSF) and a coordinate transformation. Photometric fitting is then performed on each template star in descending order of brightness, with the PSF for all other stars subtracted from the frame. When a star is found to vary significantly, it and its neighbours undergo a second iteration of fitting. The output consists of magnitudes and errors for the two colours, and six additional useful parameters (such as the χ^2 of the PSF fit and crowding information). These are used to flag questionable measurements, that arise from cosmic ray events in the CCDs, bad pixels and so on.

These photometric data are subjected to an automatic time-series analysis which uses a set of optimal filters to search for microlensing candidates and variable stars (which we have detected in abundance¹³). For each microlensing candidate a light curve is fitted, and the final selection is done automatically using criteria (for example, signal-to-noise, quality of fit, wavelength independence of the light curve and colour of the star) that were established empirically using Monte Carlo addition of fake events into real light curves.

This analysis has been done on four fields near the centre of the LMC, containing 1.8 million stars, with approximately 250 observations for each star. The candidate event reported here occurs in the light curve of a star at coordinates $\alpha = 05^{\text{h}} 14^{\text{m}} 44.5^{\text{s}}$, $\delta = -68^\circ 48' 00''$ (J2000). (A finding chart is available on request from C.A.). The star has median magnitudes $V \sim 19.6$, $R \sim 19.0$, consistent with a clump giant (metal-rich helium core burning star) in the LMC. These magnitudes are estimated using colour transformations from our filters to V and R that have been derived from observations of standard stars.

Our photometry for this star, from July 1992 to July 1993, is shown in Fig. 1, and the candidate event is shown on an expanded scale in Fig. 2, along with the colour light curve. The colour changes by <0.1 mag as it brightens and fades (the candidate 'event'). A mosaic, showing portions of some of the CCD images used, is shown in Fig. 3, with the relevant star at the centre. The integrated number of PSF photoelectrons detected above the sky background in the template image is $\sim 10^4$, for a 300 s exposure. The increase in counts during the peak is highly significant, as is clear from the figures. Also shown in Fig. 2 is a fit to the theoretical microlensing light curve (see ref. 6). The four parameters fit are (1) the baseline flux, (2) the maximum amplification $A_{\text{max}} = 6.86 \pm 0.11$, (3) the duration $t = 33.9 \pm 0.26$ d, (4) the centroid in time 433.55 ± 0.04 d. The quoted errors are formal fit errors. Using the PSF fit uncertainties as determined by the photometry program, the best-fit microlensing curve gives a χ^2 per degree of freedom of 1.6 (for 443 d.f.).

FIG. 1 The observed light curve with estimated $\pm 1\sigma$ errors. *a* Shows A_{blue} , the flux (in linear units) divided by the median observed flux, in the blue passband. *b* is the same, for the red passband.



A number of features of the candidate event are consistent with gravitational microlensing: the light curve is achromatic within measurement error, and it has the expected symmetrical shape. If this is a genuine microlensing event, the mass of the deflector can be estimated. Because the duration depends upon the lens mass, the relative velocity transverse to the line-of-sight and the distance to the lens (none of which are known), the lens mass cannot be uniquely determined from the duration. But by using a model of the mass and velocity distributions of halo dark matter, one can find the relative probability that a MACHO of mass M_{macho} gave rise to the event. Thus, if this is genuine microlensing, Fig. 9 of ref. 5 implies the most likely mass is $\sim 0.12 M_{\odot}$, with masses of $0.03 M_{\odot}$ and $0.5 M_{\odot}$ being roughly half as likely. However, this method does not properly take into account our detection efficiencies, and should be considered only a rough estimate.

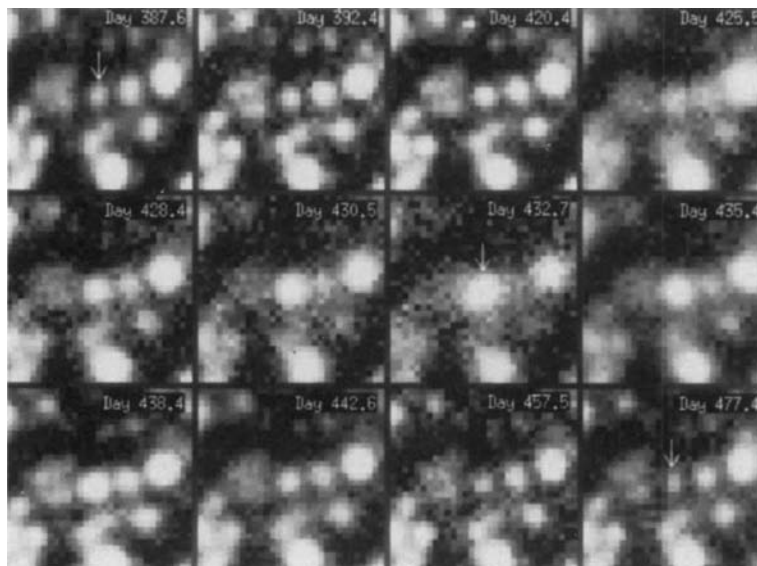
The mass range given above includes brown dwarfs and main sequence stars. Any microlensing star is very unlikely to be a red dwarf of the Galactic stellar halo, because one can show that the optical depth τ_s for microlensing by main sequence stars

of the stellar halo is very low. Even if the mass function of the stellar halo rises as steeply as $dN/dM \propto M^{-4}$, as suggested recently¹⁴ (here N is the number of stars per unit stellar mass interval), τ_s is still a few hundred times smaller than the 5×10^{-7} optical depth estimated for MACHO microlensing. The chance of finding such a stellar microlensing event among our 1.8 million stars is therefore very small.

The prospects for direct observation of a lensing object are not favourable. Even a star of $0.5 M_{\odot}$, for example, would have $V \sim 24$, and for many years would be within a small fraction of an arcsecond of the much brighter LMC star.

We emphasize that the observed stellar brightening could be due to some previously unknown source of intrinsic stellar variability. The fit discrepancy near the peak is not yet understood; a more refined analysis of the data is under way. We do not yet have a spectrum of the star. A crucial test of the hypothesis that we are seeing gravitational microlensing by MACHOs in the galactic halo will be the detection of other candidates. So far, we have analysed only $\sim 15\%$ of our first year's frames and we plan to continue observations until 1996; this should allow us

FIG. 3 Selected red CCD frames centred on the microlens candidate, showing observations before, during and after the event. The numbers on each frame indicate the days after 2 January 1992.



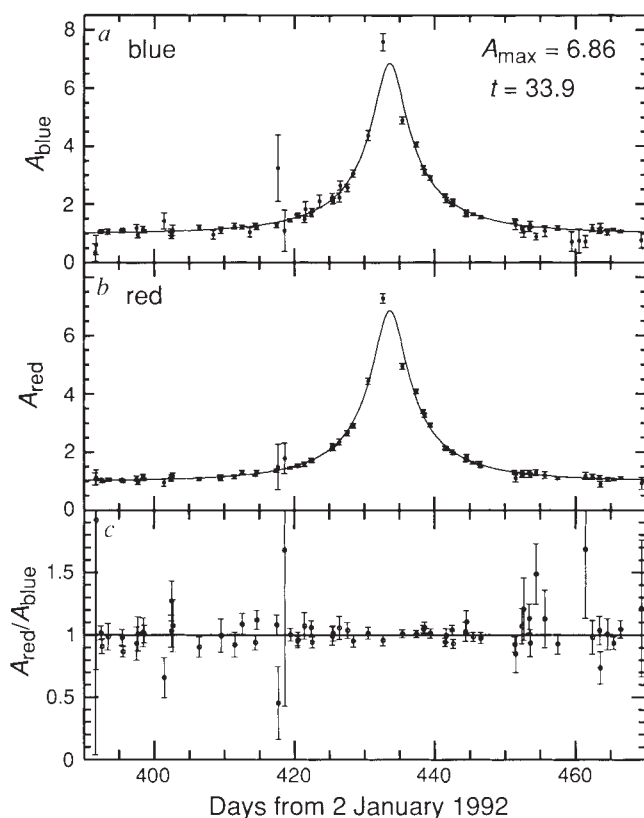


FIG. 2 As in Fig. 1, with an expanded scale around the candidate event. The smooth curve shows the best-fit theoretical microlensing model, fitted simultaneously to both c is the colour light curve, showing the ratio of red to blue flux, normalized so that the median is unity.

to determine if microlensing is really the cause. Additional events should show the theoretical distribution of maxima, and should be representative of both the colour-magnitude diagram and the spatial structure of the LMC. No repeats should be seen in any given star. (While this paper was in preparation, we were informed by J. Rich (personal communication) of the candidate events reported by the EROS collaboration. Note that the two groups use different definitions of characteristic time.)

If such candidates do result from microlensing we should be able to determine the contribution of MACHOs to the dark matter in the Galactic halo. The results presented here encourage us to believe this will happen. $\square$

Received 22 September; accepted 30 September 1993.

1. Trimble, V. A. *Rev. Astr. Astrophys.* **25**, 425–472 (1987).
2. Fich, M. & Tremaine, S. A. *Rev. Astr. Astrophys.* **29**, 409–445 (1991).
3. Primack, J. R., Seckel, D. & Sadoulet, B. A. *Rev. Nucl. Part. Sci.* **B38**, 751–807 (1988).
4. Kolb, E. W. & Turner, M. S. *The Early Universe* (Addison Wesley, New York, 1990).
5. Griest, K. *Astrophys. J.* **366**, 412–421 (1991).
6. Paczynski, B. *Astrophys. J.* **304**, 1–5 (1986).
7. Bennett, D. et al. *Ann. N.Y. Acad. Sci.* **688**, 612–618 (1993).
8. Alcock, C. et al. *Astr. Soc. Pac. Conf. Ser.* **34**, 193–202 (1992).
9. Magneville, C. *Ann. N.Y. Acad. Sci.* **688**, 619–625 (1993).
10. Udalski, A. et al. *Ann. N.Y. Acad. Sci.* **688**, 626–631 (1993).
11. Stubbs, C. W. et al. in *Charge-coupled Devices and Solid State Optical Sensors III* (ed. Blouke, M.) *Proc. of the SPIE* **1900**, 192–204 (1993).
12. Schechter, P. L., Saha, A. & Mateo, M. L. *Publ. astr. Soc. Pacif.* (in the press).
13. Cook, K. H. et al. *Bull. Am. Astr. Soc.* **24**, 1179 (1993).
14. Richer, H. B. & Fahlan, G. G. *Nature* **358**, 383–386 (1992).

ACKNOWLEDGEMENTS. We are grateful for the support given our project by the technical staff at the Mt Stromlo Observatory. Work performed at LLNL is supported by the DOE. Work performed by the Center for Particle Astrophysics on the UC campuses is supported in part by the Office of Science and Technology Centers of the NSF. Work performed at MSSSO is supported by the Bilateral Science and Technology Program of the Australian Department of Industry, Technology and Commerce. K.G. acknowledges a DOE OJL grant, and C.W.S. thanks the Sloan Foundation for their support.

Evidence for gravitational microlensing by dark objects in the Galactic halo

E. Aubourg*, P. Bareyre*, S. Bréhin*, M. Gros*, M. Lachièze-Rey*, B. Laurent*, E. Lesquoy*, C. Magneville*, A. Milsztajn*, L. Moscoso*, F. Queinnec*, J. Rich*, M. Spiro*, L. Vigroux*, S. Zylberajch*, R. Ansari†, F. Cavalier†, M. Moniez†, J.-P. Beaulieu†, R. Ferlet†, Ph. Grison†, A. Vidal-Madjar†, J. Guibert§, O. Moreau§, F. Tajahmady§, E. Maurice||, L. Prévôt|| & C. Gry¶

* DAPNIA, Centre d'Études de Saclay, 91191 Gif-sur-Yvette, France

† Laboratoire de l'Accélérateur Linéaire, Centre d'Orsay, 91405 Orsay, France

‡ Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France

§ Centre d'Analyse des Images de l'Institut National des Sciences de l'Univers, Observatoire de Paris, 61 avenue de l'Observatoire, 75014 Paris, France

|| Observatoire de Marseille, 2 place Le Verrier, 13248 Marseille 04, France

¶ Laboratoire d'Astronomie Spatiale de Marseille, Traverse du Siphon, Les Trois Lucs, 13120 Marseille, France

THE flat rotation curves of spiral galaxies, including our own, indicate that they are surrounded by unseen haloes of 'dark matter'^{1,2}. In the absence of a massive halo, stars and gas in the outer portions of a galaxy would orbit the centre more slowly, just as the outer planets in the Solar System circle the Sun more slowly than the inner ones. So far, however, there has been no direct observational evidence for the dark matter, or its characteristics. Paczyński³ suggested that dark bodies in the halo of our Galaxy can be detected when they act as gravitational 'microlenses', amplifying the light from stars in nearby galaxies. The duration of such an event depends on the mass, distance and velocity of the dark object. We have been monitoring the brightness of three million stars in the Large Magellanic Cloud for over three years, and here report the detection of two possible microlensing events. The brightening of the stars was symmetrical in time, achromatic and not repeated during the monitoring period. The timescales of the two events are about thirty days and imply that the masses of the lensing objects lie between a few hundredths and one solar mass. The number of events observed is consistent with the number expected if the halo is dominated by objects with masses in this range.

The 'EROS' (Expérience de Recherche d'Objets Sombres) collaboration is searching for microlensing events using the European Southern Observatory at La Silla, Chile^{4,5}. We have two complementary programmes. The first uses $5^\circ \times 5^\circ$ Schmidt plates of the Large Magellanic Cloud (LMC) that allow us to monitor about eight million stars with a sampling rate of no more than two measurements per night. This makes the programme primarily sensitive to lens masses in the range $10^{-4} M_\odot < M < 1 M_\odot$ (where $M_\odot$ is the solar mass), corresponding to mean lensing durations in the range $1 \text{ d} < \tau < 100 \text{ d}$. The probability that a given star in the LMC is amplified by more than 0.3 magnitudes at a given time is calculated to be $\sim 0.5 \times 10^{-6}$ (refs 3, 6). For a deflector of mass M the typical timescale for the amplification is $\tau = 70 \sqrt{M/M_\odot} \text{ d}$. The light curve of such an event should be symmetric in time, achromatic, and the event should not be repeated. Over the period 1990–93, a total of 304 Schmidt plates of the LMC were taken for us at La Silla with red or blue filters. Exposure times were typically

one hour, permitting us to monitor stars down to the twentieth magnitude with a mean photometric precision of about 15% (r.m.s.). The transparency of the plates is digitized in $10\ \mu\text{m}$ ($0.67\ \text{arcsec}$) steps by the 'MAMA' (Machine Automatique a Mesurer pour l'Astronomie) at the Observatoire de Paris⁷. The relation between transparency and star luminosity has been established using charge-coupled device (CCD) images scattered through the Schmidt-plate field.

The second programme uses a CCD camera consisting of a mosaic of sixteen 579×400 pixel Thomson THX 31157 CCDs covering about $1^\circ \times 0.4^\circ$. It is mounted on a 40-cm reflector (f/10) refurbished with the help of the Observatoire de Haute Provence. We have used this to observe one field in the bar of the LMC from December 1991 to March 1992 and from August 1992 to March 1993. As of March 1993, a total of 8,100 exposures had been taken with red and blue filters. About 100,000 stars are seen on each image, with a mean photometric precision of about 6%. Compared to the Schmidt-plate programme, the number of stars is a factor 80 smaller but the rapid sampling time (an image pair every 22 minutes) makes the CCD programme sensitive to deflector masses in the range $10^{-7} M_\odot < M < 10^{-3} M_\odot$, corresponding to event durations in the range $1\ \text{h} < \tau < 3\ \text{d}$.

After preliminary processing (digitizing the Schmidt plates and flat-fielding the CCD images), the data reduction for both programmes follows basically the same procedure. First, one reference image for each colour was constructed by combining ten plates or 50 CCD images taken with good atmospheric conditions. We used a star finding algorithm to establish a star catalogue for each reference image. Next, each image is aligned with the reference using bright, isolated stars. The positions of the stars on the reference image then serve as input to a photometric fitting programme to determine the luminosity of each catalogue star on the new image. The image is then aligned 'photometrically' with the reference by requiring that the mean luminosity of stars in a given luminosity band equal the mean luminosity in the catalogue. (The small number of intrinsically variable stars in the catalogue does not affect this procedure.)

TABLE 1 Characteristics of the two microlensing candidates

	Candidate 1	Candidate 2
Coordinates of star (J2000)	$\alpha = 5\ \text{h}\ 26\ \text{m}\ 36\ \text{s}$ $\delta = -70^\circ\ 57'\ 37''$	$\alpha = 5\ \text{h}\ 06\ \text{m}\ 05\ \text{s}$ $\delta = -65^\circ\ 58'\ 34''$
b magnitude	19.3 ± 0.2	19.3 ± 0.2
b-r	0.3 ± 0.2	0.0 ± 0.2
Date of maximum amplification	1 February 1992	29 December 1990
Event duration (τ in days)	27 ± 2	30 ± 3
Maximum amplification in magnitudes (blue filter)	1.0 ± 0.1	1.1 ± 0.2
Maximum amplification in magnitudes (red filter)	1.0 ± 0.1	1.3 ± 0.2
Maximum amplification in magnitudes (combined fit)	1.0 ± 0.1	1.2 ± 0.2
χ^2 (combined fit)	192 for 248 d.o.f.	167 for 131 d.o.f.

Errors in the magnitudes include fitting errors as well as systematic uncertainties in the magnitude-plate transparency relation. The χ^2 were calculated using the estimated errors which are known only to an estimated precision of 20%. The timescale τ is the time taken by the dark object to cross an angle corresponding to one 'Einstein radius' (ref. 3). This definition differs from the traditional one which is the time the dark object remains within the Einstein ring. The traditional definition is meaningless for lensing events with a minimum impact parameter greater than the Einstein radius.

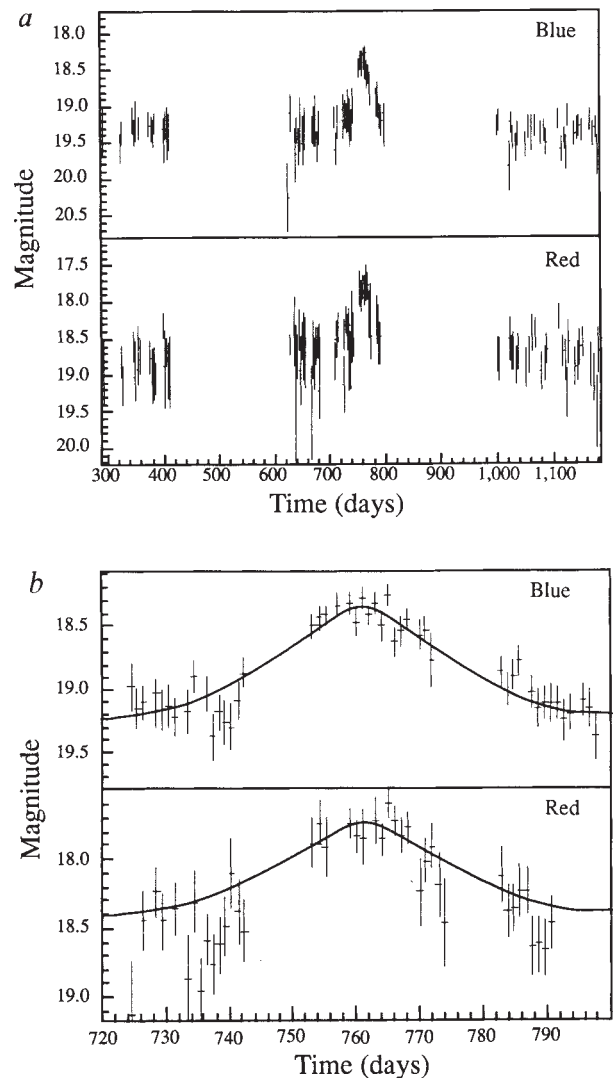


FIG. 1 a, The measured magnitudes for candidate 1 as a function of time. The time is counted from 1 January 1990. The error bars correspond to the estimated 1σ errors. b, The light-curve of candidate 1 on an expanded scale. The curve shows the best fit for the microlensing hypothesis. The parameters of the best fit are shown in Table 1.

Successive images then add one point to the blue or red light curve of each star in the catalogue.

After data reduction, each light curve is tested for the presence of time variations using a variety of algorithms. For microlensing-like events, we use a simple algorithm that scans curves for sequences of measurements that are significantly above the mean value. The light curve is selected as a microlensing candidate if it exhibits one and only one such sequence simultaneously in both colours. The precise value of the threshold for acceptance is chosen using estimates of measurement errors so that random fluctuations of intrinsically stable stars are not accepted. The photometric errors are estimated as a function of magnitude for each image but may still vary from one star to another by 20% according to the star's environment. We estimate the efficiency of the cuts to accept real microlensing events using Monte Carlo-generated lensing events, produced with the observed photometric resolution and observing sequence. We superimposed the Monte Carlo signature of microlensing events onto both observed (flat) light curves and simulated light curves. These curves were then subjected to the same algorithms as the real data.

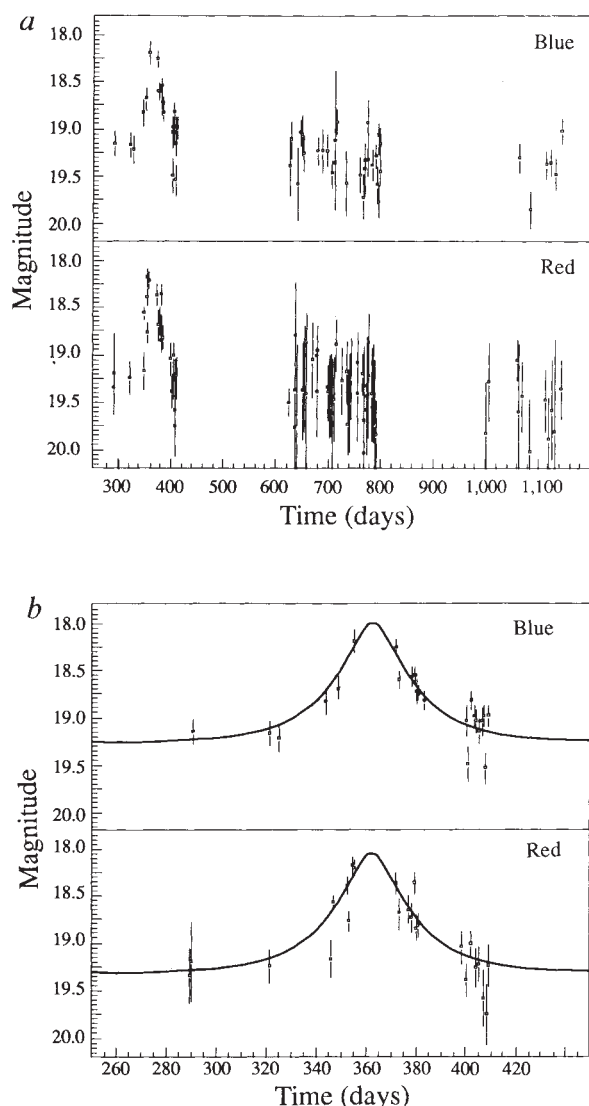


FIG. 2 As Fig. 1, for candidate 2.

For simulated events with peak amplifications >1.34 inserted into the data during the period of the observing seasons, the efficiency ranges from $\sim 25\%$ for events with timescales of 6 days to 50% for events with timescales of 30 days. These numbers differ from 100% because of the sampling period and the photometric resolution for faint stars.

Measured curves passing the above selection criteria were then inspected visually and subjected to further analysis to determine their compatibility with the microlensing hypothesis. At this stage, we found it necessary to eliminate only those light curves that exhibit variations on a timescale comparable to that of the total observing period. Such events cannot be tested for the presence of subsequent variations. Remaining events are fitted for the theoretical microlensing light curve. The parameters of the fit are the off-lensing luminosity, the maximum amplification, the time of maximum amplification, and the timescale of the microlensing. The light curves of the two colours are fitted separately (to test for wavelength independence) and then simultaneously.

We found no candidates in an analysis performed on the 1991–92 CCD data (20% of the total data). For this data, we expected about three candidates if the halo is entirely comprised of dark objects in the range 10^{-7} to $10^{-5} M_{\odot}$.

A preliminary analysis of 40% of the Schmidt-plate data has

revealed two events that are consistent with the microlensing hypothesis. The light curves are shown in Figs 1 and 2, with the event characteristics listed in Table 1. They are the only curves so far analysed that show one significant amplification event with no further variations. (No curves have been found that show two or more examples of microlensing-like behaviour.) The curves are consistent with the theoretical curve; χ^2 s are good within the estimated 20% uncertainty in the photometric errors. The events are wavelength independent at the 10% level, that is within errors. The amplification is near the median amplitude expected ($\delta m = 1.0$) for detectable events with these time-scales. Neighbouring stars show no variations over the whole observing period. The off-lensing magnitudes of the two stars are near the average for stars in our catalogue. Candidate 2 is on the main sequence while candidate 1 is between the main sequence and giant branch of the colour-magnitude diagram. Only the observation of further events will tell us if there is an accumulation of events in a given region of the diagram indicating variable-star phenomena.

If the events are interpreted as arising from microlensing, the lensing objects would have a mass between a few $\times 10^{-2}$ and $1 M_{\odot}$. This range is based on simulations with the standard isothermal halo yielding the observed flat rotation curve out to the LMC (refs 2, 6). The range is wide because, for fixed mass, the lensing time varies due to the uncertain distance and speed of the lensing object. The allowed masses include those expected for brown dwarfs and dim main sequence stars. If they are located in the Galactic halo, they would not be detectable. If the halo consists entirely of such objects, we estimate that we would have observed about six events if the dominant mass is $10^{-2} M_{\odot}$ and about one event if it is $1 M_{\odot}$, in agreement with the observed number.

Obviously, this interpretation of the events must be confirmed by further observations of the stars in question and, especially, by the discovery of further events, which would permit us to study their statistical properties. In particular, the distribution of amplification magnitudes must be shown to be consistent with the microlensing hypothesis and the microlensed stars must be distributed throughout the observed colour-magnitude diagram and throughout the LMC. Until this is done, it is not possible to rule out the possibility that we are dealing with a new type of variable star. During the preparation of this paper we learned that a similar microlensing event has been observed by the 'MACHO' collaboration (C. Alcock, personal communication).

The EROS collaboration is continuing to collect and analyse data and it is expected that further results will be presented within a year's time.

Note added in proof: We have located in our data the star involved in the candidate microlensing presented by the MACHO collaboration (C. Alcock *et al.*, this issue). Because the star is very faint in the blue band, we have reliable measurements only in the red. (This fact explains the rejection of the event in our analysis.) We confirm in this colour the existence and characteristics of the MACHO event. No other significant luminosity variations are seen in the 1990–91 and 1991–92 seasons, reinforcing the microlensing interpretation of the event. $\square$

Received 22 September; accepted 30 September 1993.

1. Trimble, V. A. *Rev. Astr. Astrophys.* **25**, 425–472 (1987).
2. Primack, J. R., Seckel, D. & Sadoulet, B. A. *Rev. Nucl. Part. Sci.* **38**, 751–807 (1988).
3. Paczyński, B. *Astrophys. J.* **304**, 1–5 (1986).
4. Aubourg, E. *et al. Messenger* **72**, 20–27 (1993).
5. Aubourg, E. thesis, Univ. Paris (1992).
6. Griest, K. *et al. Astrophys. J.* **372**, L79–82 (1991).
7. Berger, J. *et al. Astr. Astrophys. Suppl. Ser.* **87**, 389 (1991).
8. Alcock, C. *et al. Nature* **365**, 621–623 (1993).

ACKNOWLEDGEMENTS. We thank C. Alcock for discussions at the beginning of this project. We thank A. Bijaoui, Ph. Veron, the staff at the Observatoire de Haute Provence, the ESO staff at La Silla, and the technical staff of the collaborating laboratories for their advice and help. We thank D. Bennett of the MACHO collaboration for help and interesting discussions. This work is based on observations at the European Southern Observatory, La Silla, Chile, and is funded by DSM-CEA, IN2P3-CNRS, INSU-CNRS, with support from ESO.

Solving the mystery of the X-ray periodicity in the Seyfert galaxy NGC6814

Greg M. Madejski^{*†}, Chris Done^{*‡},
T. Jane Turner^{*†}, Richard F. Mushotzky^{*},
Peter Serlemitsos^{*}, Fabrizio Fiore[§],
Marek Sikora^{¶¶} & Mitchell C. Begelman^{¶¶}

^{*} Laboratory for High Energy Astrophysics, Code 666, NASA, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

[†] Universities Space Research Association

[§] Center for Astrophysics, Cambridge, Massachusetts 02138, USA

[¶] Copernicus Astronomical Center, Bartycka 18, 00-716 Warsaw, Poland

^{¶¶} Joint Institute for Laboratory Astrophysics, University of Colorado, Boulder, Colorado 80309, USA

[‡] Department of Astrophysical, Planetary, and Atmospheric Sciences, University of Colorado, Boulder, Colorado 80309, USA

ACTIVE galaxies, of which Seyfert galaxies are a subgroup, are thought to be powered by the accretion of gas onto a massive black hole at the galaxy centre; the X-rays emitted by active galactic nuclei arise from the heated, infalling gas. The observed rapid variability of this X-ray emission requires a compact source, providing additional support for the black-hole hypothesis. The Seyfert galaxy NGC6814 has been thought to be unique among active galaxies, in exhibiting not simply variability but a periodicity of $\sim 12,100$ s in its X-ray luminosity¹⁻⁴. Many exotic theories have been proposed to explain this periodicity, including gravitational lensing of hotspots on the accretion disk by the central black hole⁵⁻⁷ or the effects of a captured star orbiting the black hole^{4,8-10}. Here we show, using data from the Rosat X-ray telescope, that although the 12,100 s period is indeed real and stable, it is not associated with NGC6814. Instead, the periodic emission arises from another source, 37 arcmin away from NGC6814, which is probably an object in our Galaxy—for example, a white dwarf accreting gas (from a giant companion) onto its poles.

NGC6814 was observed by the Rosat X-ray telescope with the Position Sensitive Proportional Counter PSPC-B from 31 March–2 April 1993. Figure 1 shows the X-ray image over the 2-degree diameter field-of-view in the full 0.1–2.4 keV bandpass of the instrument. Many sources can be seen in the field of view, with NGC6814 being the bright source at the center. Its derived X-ray position, RA (2000) = 19 h 42 min 40.4 s, dec. (2000) = $-10^\circ 19' 17.5''$ (with a nominal uncertainty of $\sim 10''$) is identical to the optical position of the NGC6814 nucleus¹¹.

Another bright point source can be seen ~ 37 arcmin away, although the image is strongly distorted from point-like because of the spherical aberration of the instrument at large off-axis angles. The position derived from the Rosat image is RA (2000) = 19 h 40 min 13 s, dec. (2000) = $-10^\circ 25' 30''$, where we estimate the position error to be ~ 30 arcsec, larger than the nominal value because this source is off-axis and partially obscured by the PSPC window-supporting rib. Our examination of the light curves of both sources suggested that the off-axis one is more variable.

We tested the time series from both sources for periodicity in the 10,000–15,000 range by repeatedly folding the light curve and measuring the significance of deviations from a constant by the L-statistic (related to χ^2 but following an F distribution^{3,12,13}). Figure 2a and b show the L-statistic for NGC6814 and the second source, respectively. The 12,000 periodicity is clearly not associated with the Seyfert galaxy but with the off-axis source; we show its light curve folded on the best

fitting period of 12,142 s in Fig. 3. The error in this derived period is determined primarily by the number of periods sampled in the observation and, as the Rosat light-curve is very similar in length and sampling to the Ginga data, we estimate that the uncertainty in the period is also similar, that is ~ 20 s (ref. 4). We note that this periodicity cannot be due to the wobble of the spacecraft, which occurs on a timescale of only ~ 400 s, nor the orbital period of Rosat, which is ~ 96 min (5,760 s).

The soft X-ray spectrum of this second, off-axis, source can be modelled as a black body plus a power law, modified by photoelectric absorption from intervening material in our Galaxy. The model gives a black body temperature of $k_B T_{bb} = 63 \pm 11$ eV and a power law photon index of $\Gamma = 1.1 \pm 0.25$ ($\chi^2 = 12.8/20$ energy channels). The average 0.2–2 keV flux from the source is $\sim 6 \times 10^{-12}$ erg cm⁻² s⁻¹ (uncertain by $\sim 20\%$ because of partial obscuration by the PSPC entrance window-support ribs) with the blackbody component contributing $\sim 45\%$ of this flux, and the power law component responsible for the remainder. The PSPC spectrum of NGC6814, on the other hand, can be described by a power law with $\Gamma = 1.79 (+0.44, -0.33)$, again modified by absorption from material in the line of sight through our Galaxy¹⁴. The total inferred 0.2–2 keV flux from the Seyfert is 7.1×10^{-13} erg cm⁻² s⁻¹.

The low Galactic latitude of the region ($b \sim -16^\circ$) strongly suggests a source in our Galaxy. Of the known classes of X-ray emitting objects, only low-mass X-ray binaries (LMXBs) and cataclysmic variables (CVs) have the correct range of orbital periods¹⁵ to explain the observed periodicity. The ~ 60 eV black body temperature suggests a white dwarf primary, as the neutron star primary in a LMXB would have a higher temperature, ~ 1 keV, because of the deeper gravitational potential well. The fact that the 2–20 keV Ginga spectrum (dominated by the periodic source¹³) is hard^{16,17} (flat) can further discriminate among the different types of CV systems¹⁸. Typically, only white dwarf binaries with strong magnetic fields that disrupt disk formation and channel the accretion flow onto the magnetic poles give rise to spectra that are flat in the 2–20 keV band (although quiescent disk-fed white dwarf systems can also show flat 2–20 keV spectra¹⁸). There are two types of magnetic cataclysmic variables; the AM Herculis systems, where the white dwarf spin is phase-locked with the orbital period, and the DQ Herculis systems, where the phase-locking is not complete. As the spin period of the white dwarf is not clearly seen in either the Exosat or Ginga data, we favour identification with an AM Her system.

We have shown that the X-ray emission with $\sim 12,100$ s periodicity, previously attributed to the Seyfert 1 galaxy NGC6814, arises from an unrelated object that is located ~ 37 arcmin from the nucleus of NGC6814. Considering the limited spatial resolution of collimated proportional counter instruments such as HEAO-A2, Ginga LAC, and Exosat Medium Energy Counter, the original misidentification is not entirely surprising, and a sensitive, large field-of-view imaging instrument such as the Rosat PSPC was necessary for the correct identification. Other imaging instruments either had too small a field-of-view (Einstein Observatory Imaging Proportional Counter), or too high a background count and off-axis distortion (Exosat Low Energy Telescope) to show an obvious contaminating source.

This finding eliminates a lot of difficulties with the modelling of the X-ray emission from NGC6814 as most, if not all, of the 2–20 keV flux comes from the secondary source rather than the Seyfert nucleus itself. The flat X-ray spectrum, with $\Gamma \sim 1.4$ (ref. 16,17) was quite unusual for an AGN, and $\Gamma \sim 1.8$ is more typical. The inferred periodicity required fine-tuning of the source geometry³⁻¹⁰, and was unique among active galaxies, whereas it is expected from a Galactic accreting binary. The rapid drop of flux observed in the HEAO A-2 (ref. 19) and Ginga data^{3,4,6,20} implied an extremely compact source, and the combination of small size and large inferred luminosity (assuming the X-rays came from NGC6814) required a supermassive black hole^{3,4,16,20}, whereas short timescales are easily explained in the much smaller

[†] Present address: Physics Department, Leicester University, Leicester, UK.

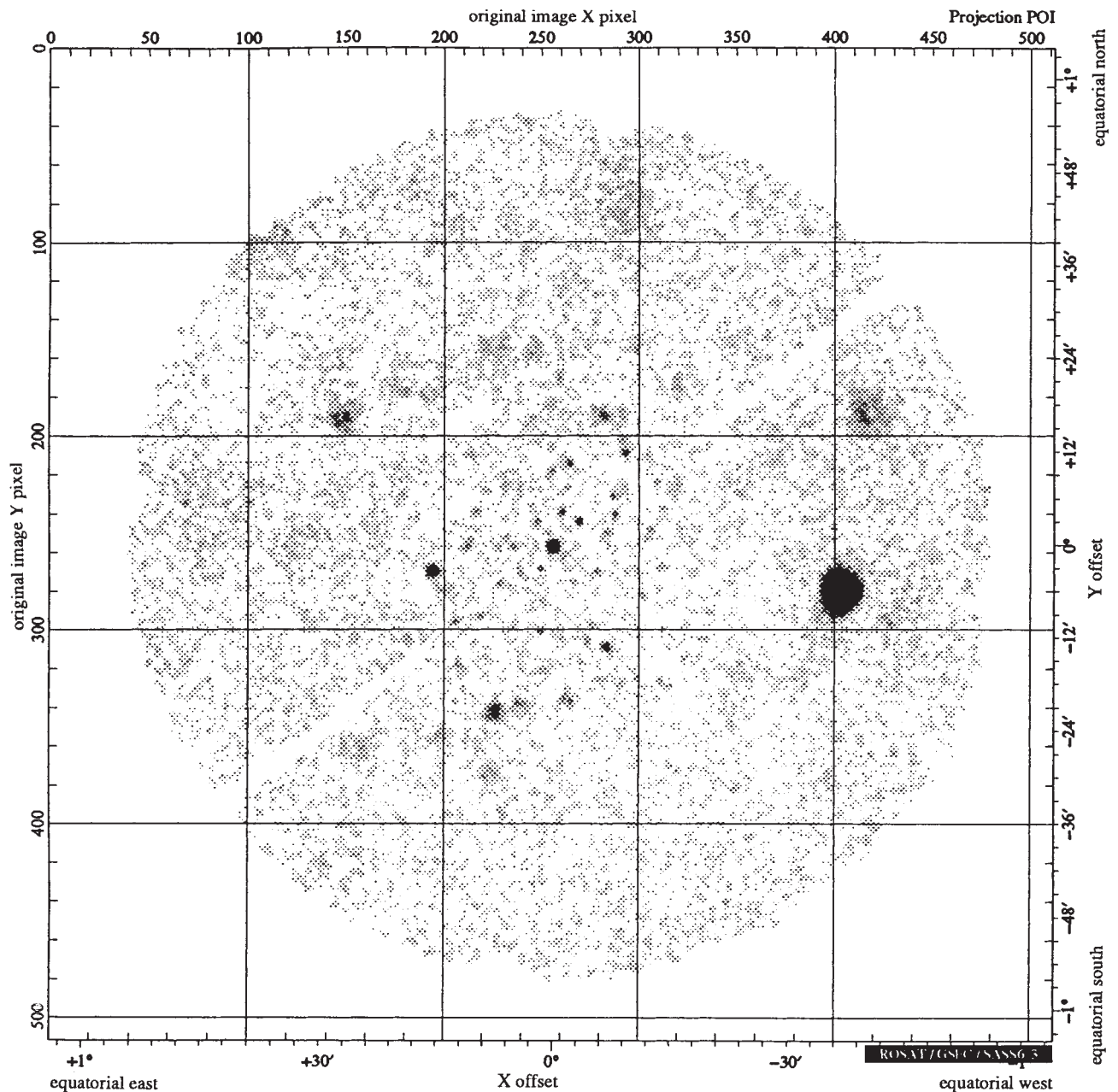


FIG. 1 The Rosat PSPC image of the NGC6814 field, centred at RA (2000) = 19 h 42 min 40.4 s, dec. (2000) = $-10^{\circ} 10' 25''$. The observation was made over 180,000 s of running time, yielding 38,000 s of

good data. The on-axis source is the Seyfert galaxy NGC6814, while the very bright source, 37' west, with the image smeared by the degraded off-axis point spread function, is the origin of the periodic X-ray emission.

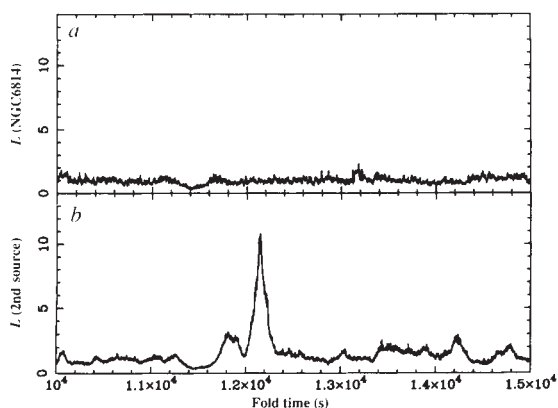


FIG. 2 Results of L-statistic tests for the periodicity of NGC6814 (a) and off-axis source (b) over the range 10,000 to 15,000 s. The off-axis source shows a clear period at 12,142 s.

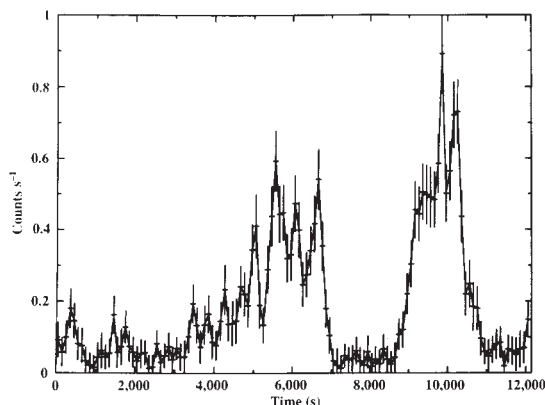


FIG. 3 The light curve of the off-axis source, folded on a 12,142-s period. A structure similar to that seen in the Ginga data^{3,20} is apparent.

size-scales of a Galactic source. The strong and rapidly variable iron K α emission line seen in the Ginga data^{16,17} was cited as another piece of evidence in support of a supermassive black hole in NGC6814. Its strength and variability, thought to arise from the X-ray illuminated accretion flow in the active nucleus, required a substantial amount of ionized material close to the black hole¹⁷. This produces a line at 6.7 keV, so a strong gravitational redshift was required^{16,17} in order to be consistent with the observed line energy of ~ 6.4 keV. Large and variable emission lines from ionized material, however, are common in AM Her objects²² and the apparent discrepancy of the line energy likely arises from the large distortion induced in the Ginga spectrum because the source is so far off axis.

Even though our result eliminates an important piece of observational evidence for black holes as the central engine in active galactic nuclei, we stress that the small central source sizes inferred from the X-ray variability of Seyfert galaxies as a class are unaffected by this finding. These sources often exhibit non-periodic, large-amplitude variability on time scales of thousands of seconds²³, and the removal of NGC6814 does not substantially change the strong support that the rapid X-ray variability provides for the black hole model of active galactic nuclei²⁴.

Note added in proof: After submitting this manuscript, we optically identified the periodic (off-axis) source with a 16 mag variable star, exhibiting strong He II 4,686 line emission²⁵. The spectra also show phase-dependent, broad-emission features which we associate with cyclotron emission, thus confirming our suggestion that the source is a magnetic CV system. $\square$

Received 20 July; accepted 31 August 1993.

- Mittaz, J. P. D. & Branduardi-Raymont, G. *Mon. Not. R. astr. Soc.* **238**, 1029–1046 (1989).
- Fiore, F., Massaro, E. & Barone, P. *Astr. Astrophys.* **261**, 405–414 (1992).
- Done, C. *et al. Astrophys. J.* **400**, 138–152 (1992).
- King, A. & Done, C. *Mon. Not. R. astr. Soc.* (in the press).
- Abramowicz, M., Bao, G., Lanza, A. & Zhang, X.-H. *Astr. Astrophys.* **245**, 454–456 (1991).
- Abramowicz, M. A. *et al. in The Physics of Active Galactic Nuclei* (eds Wagner, S. & Dushli, W. J. 61–66 (Springer, Heidelberg, 1992).
- Abramowicz, M., Bao, G., Karas, V. & Lanza, A. *Astr. Astrophys.* **272**, 400–406 (1993).
- Syer, D., Clarke, C. J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **250**, 505–512 (1991).
- Sikora, M. & Begelman, M. C. *Nature* **356**, 224–225 (1992).
- Chakrabarti, S. K. *Astrophys. J.* **411**, 610–613 (1993).
- Clements, E. D. *Mon. Not. R. astr. Soc.* **197**, 829–834 (1981).
- Davies, S. R. *Mon. Not. R. astr. Soc.* **244**, 93–95 (1990).
- Davies, S. R. *Mon. Not. R. astr. Soc.* **251**, 64p–65p (1991).
- Elvis, M., Lockman, F. J. & Wilkes, B. *Astr. J.* **97**, 777–782 (1989).
- Patterson, J. *Astrophys. J. Suppl. Ser.* **54**, 443–493 (1984).
- Kunieda, H. *et al. Nature* **345**, 786–788 (1990).
- Turner, T. J., Done, C., Mushotzky, R. F., Madejski, G. M. & Kunieda, H. *Astrophys. J.* **391**, 102–110 (1992).
- Cordova, F. *in X-ray Binaries* (eds Lewin, W. H. G., van Paradijs, J. & van der Heuvel, E. P. J.) (in the press).
- Tennant, A., Mushotzky, R. F., Boldt, E. A. & Swank, J. H. *Astrophys. J.* **251**, 15–25 (1981).
- Leighly, K., Kunieda, H., Tsusaka, Y., Awaki, H. & Tsuruta, S. *Astrophys. J.* (in the press).
- Hayakawa, S. *Nature* **351**, 214–215 (1991).

- Ishida, M. thesis, Univ. Tokyo (1991).
- McHardy, I. in *Two Topics in X-ray Astronomy*, Proc. 23rd ESLAB Symposium (eds White, N., Hunt, J. & Battrick, B. 1111–1124 (European Space Agency, Paris, 1989).
- Fabian, A. C. in *Frontiers of X-ray Astronomy*, (eds Tanaka, Y. & Koyama, K.) 603–610 (Universal Academy, Tokyo, 1992).
- Rosen, S., Done, C., Watson, M. & Madejski, G. *IAU Circ. No. 5850*.

ACKNOWLEDGEMENTS. This research has been supported by NASA.

Efficient light-emitting diodes based on polymers with high electron affinities

N. C. Greenham*, S. C. Moratti†, D. D. C. Bradley*, R. H. Friend* & A. B. Holmes†

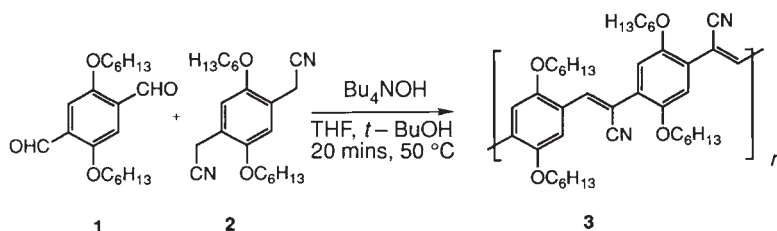
* Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

† University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

CONJUGATED polymers have been incorporated as active materials into several kinds of electronic device, such as diodes, transistors¹ and light-emitting diodes². The first polymer light-emitting diodes were based on poly(*p*-phenylene vinylene) (PPV), which is robust and has a readily processible precursor polymer. Electroluminescence in this material is achieved by injection of electrons into the conduction band and holes into the valence band, which capture one another with emission of visible radiation. Efficient injection of electrons has previously required the use of metal electrodes with low work functions, primarily calcium; but this reactive metal presents problems for device stability. Here we report the fabrication of electroluminescent devices using a new family of processible poly(cyanoterephthalylidene)s. As the lowest unoccupied orbitals of these polymers (from which the conduction band is formed) lie at lower energies than those of PPV, electrodes made from stable metals such as aluminium can be used for electron injection. For hole injection, we use indium tin oxide coated with a PPV layer; this helps to localize charge at the interface between the PPV and the new polymer, increasing the efficiency of recombination. In this way, we are able to achieve high internal efficiencies (photons emitted per electrons injected) of up to 4% in these devices.

Since our first report of electroluminescence in poly(*p*-phenylenevinylene), (PPV)², a variety of conjugated polymers^{3–7} have been shown to exhibit electroluminescence^{8,9}. A typical polymer light-emitting diode (LED) comprises a thin film of polymer sandwiched between metal electrodes, one of which is semi-transparent. Under an applied bias, oppositely charged carriers are injected from the opposing contacts and are swept through the device by the electric field. These carriers may then capture one another within the device to form excitons, and the singlets amongst them may then decay radiatively, giving out light at a wavelength characteristic of the energy gap of the polymer. In order to achieve high electroluminescence efficiency, it is necessary to balance the rates of injection of electrons and holes from opposite contacts into the device. For the conjugated polymers investigated so far, electron injection has proved more difficult than hole injection, and it has been necessary to use metals with low work functions (such as calcium) as the electron-injecting contact in order to achieve good efficiencies³. Calcium is highly susceptible to atmospheric degradation, and is difficult to encapsulate, so it would clearly be advantageous to be able to use metals with higher work functions (such as aluminium) which possess greater stability. In the present work we have designed polymers with increased electron affinity to reduce the barrier to electron injection. Another strategy to improve device efficiency is to use additional organic charge-transporting layers between the emissive layer and one or both of the electrodes.

FIG. 1 Synthesis of the cyano-substituted polymer 3. Equimolar quantities of compounds 1 and 2 were treated with 5 mol% $(C_4H_9)_4NOH$ in tetrahydrofuran and *t*-butanol (*t*BuOH) (1:1) at 50 °C for 20 min, and the resulting red precipitate was filtered and washed with methanol. 1H NMR: ($CDCl_3$, 250 MHz) δ 0.7–2.0 (22H, m), 4.09 (4H, m), 7.16 (1H, s), 7.93 (1H, s), 8.10 (1H, s), where δ indicates proton chemical shift against tetramethylsilane as internal standard ($\delta=0$), m and s indicate multiplet and singlet respectively. Gel permeation chromatography against polystyrene standards indicates a number-average molecular weight of 4,000.



The aim is to introduce energy barriers to electrons or holes at the heterojunction between the layers, and thus to confine charge carriers within the device. The confined charge at the interfaces acts to redistribute the electric field within the device so as to improve the balance of electron and hole injection. The emission is also moved away from the metal electrodes which may act as quenching sites for singlet excitons. This technique is well-established in molecular organic LEDs^{10,11} and we have shown recently that it is possible to achieve improved efficiency in a PPV LED by incorporating an electron-transporting layer, formed as a blend of a molecular semiconductor in a polymeric matrix, between the PPV and the electron-injecting contact^{12,13}. Other groups have demonstrated the use of hole-transporting layers with conjugated polymers^{14,15}. We report here the use of PPV as a hole-transporting layer together with our new polymers to achieve greatly increased efficiencies.

Insoluble cyano-substituted polymers based on poly(arylenevinylene)s have been reported by several groups^{16,17}. We have designed a family of soluble highly fluorescent poly(cyano-terephthalylidene)s in which each arylene fragment carries solubilizing long-chain alkoxy substituents. The polymer (3) was synthesized by Knoevenagel condensation polymerization of the 2,5-bis(hexyloxy)terephthalaldehyde (1) with the 2,5-bis(hexyloxy)benzene-1,4-diacetonitrile (2) as shown in Fig. 1. Careful control of reaction conditions is required to avoid Michael addition which may lead to inferior polymers.

Optical absorption and emission (photoluminescence) spectra of the polymer 3 are shown in Fig. 2. Thin films (100–300 nm)

were formed by spin-coating from chloroform solution onto glass substrates coated with indium tin oxide. Other devices were formed by spin-coating the tetrahydrothiophenium leaving-group precursor to PPV¹⁸ onto the indium tin oxide substrate. The precursor was then converted to PPV by heating at 220 °C for 12 h under vacuum to give a final PPV layer of thickness in the range 100–270 nm before depositing 3. Electron-injecting contacts (area 10 mm²) of calcium, aluminium or gold were formed by vacuum evaporation. Current–voltage characteristics of the devices were measured while simultaneously measuring the light output using a calibrated silicon photodiode. Internal quantum efficiencies were calculated, taking into account refraction at the polymer–glass–air interfaces, and the electroluminescence emission spectra were measured.

Single-layer devices of polymer 3 showed uniform red emission under forward bias, with the same spectrum as the photoluminescence. Fields of $\sim 1.2 \times 10^6$ V cm⁻¹ were required to obtain current densities of 5 mA cm⁻². Similar internal quantum efficiencies, up to 0.2%, were obtained with both calcium and aluminium contacts. For other conjugated polymers, such as PPV⁸, at least a tenfold reduction in efficiency is found in changing the electron-injecting contact from calcium to aluminium, which has a work function of the order of 1 eV greater than that of calcium. These results indicate that electron injection is greatly improved by the introduction of cyano groups at some of the ethylidene linkages in poly(arylenevinylene)s, with the effect that device efficiency is now determined by processes at the indium tin oxide electrode.

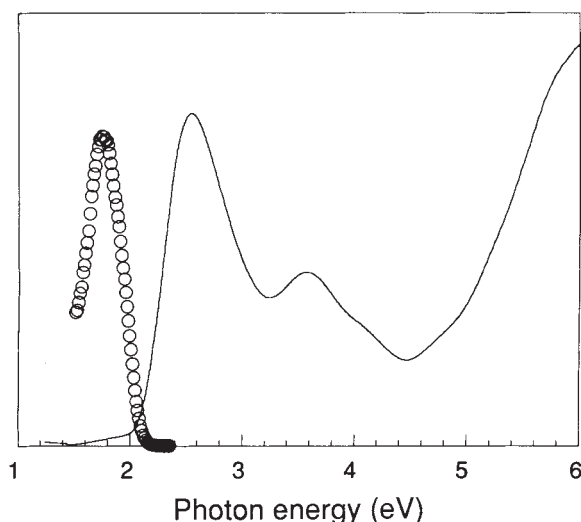


FIG. 2 Absorption (solid line) and emission (circles) spectra of a thin film of polymer 3. Emission spectra were measured under laser excitation at 457 nm, using a monochromator and photomultiplier tube whose spectral response had been calibrated using a tungsten lamp of known spectral output. The peak emission wavelength is 710 nm. (Vertical axis is in arbitrary units, linear scale.)

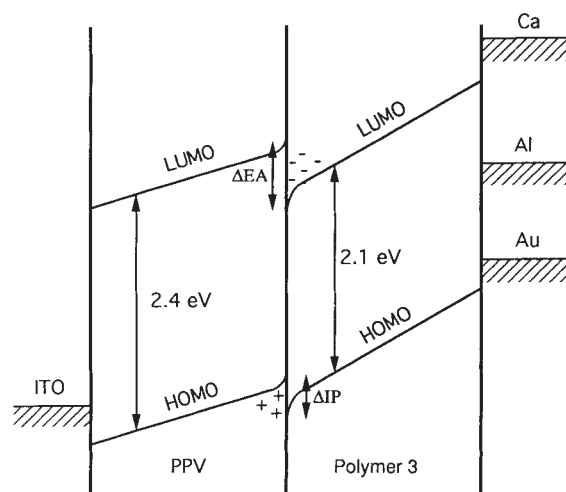


FIG. 3 Schematic energy-level diagram for a bilayer device under forward bias. The absolute energies of many of the levels are not known accurately, but the diagram shows the relative positions of the HOMO and LUMO levels in the polymers, and the Fermi levels of the various possible metal contacts. The differences in electron affinity (δEA) and ionization potential (δIP) between PPV and polymer 3 are shown. ITO, indium tin oxide.

Polymer bilayer devices using PPV as a hole-transporting layer also have uniform red electroluminescence. The spectrum was again similar to the photoluminescence spectrum of 3, with no evidence of emission from the PPV layer. The introduction of the PPV layer causes a significant reduction in the field (drive voltage/total polymer layer thickness) required to drive the device. A current density of 5 mA cm^{-2} requires only a field of $\sim 4 \times 10^5 \text{ V cm}^{-1}$ with aluminium or calcium electrodes, comparable to the best of the sublimed molecular film devices^{10,11}. Outputs of $5.2 \text{ mW sr}^{-1} \text{ A}^{-1}$ were obtained normal to the plane of the device, giving a brightness of $2.6 \text{ W sr}^{-1} \text{ m}^{-2}$ at a current density of 50 mA cm^{-2} . The emission was roughly lambertian over the entire forward hemisphere. This corresponds to an internal quantum efficiency of 4%. Again, comparable efficiencies were obtained with aluminium and calcium electrodes. With a gold electrode, typical efficiencies of 1% were obtained, but increased fields were required (approximately $7 \times 10^5 \text{ V cm}^{-1}$ for 5 mA cm^{-2}). The reduction in drive field compared with the single-layer devices indicates that charge injection is improved in the bilayer structure, and the increase in efficiency is consistent with significant charge confinement at the interface between the PPV and polymer 3 layers.

Data from differential pulse polarography on insoluble poly(cyanoterephthalylidene)s¹⁹ indicates that the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) levels are lowered by 0.6 and 0.9 eV, respectively, compared with those of PPV. Preliminary cyclic voltammetry results for polymer 3 show similar trends. These offsets are large enough to present a significant barrier at the heterojunction to both electrons and holes at room temperature, leading to charge confinement. A schematic energy level diagram is shown in Fig. 3. By confining charges at the interface, a large field is established which promotes tunnelling through the barriers. We believe that emission occurs either when a hole tunnels through the barrier and forms an exciton directly in polymer 3, or when an electron tunnels into the PPV and forms an exciton which then diffuses into the lower-energy polymer 3 and decays there. Exciton generation thus takes place very close to the heterojunction, so that emission is seen only from the lower-gap polymer, which is, in this case, the cyano-substituted PPV. We consider that device efficiency, and the drive voltage required to establish the forward current, are determined primarily by the

characteristics of the polymer-polymer heterojunction rather than the two electrode-polymer contacts.

The synthetic route to polymer 3 lends itself to considerable variation. By changing the monomer units, we have prepared a range of polymers exhibiting improved electroluminescence efficiencies with a range of emission spectra. An example of an orange/yellow-emitting polymer (4) is shown in Fig. 4. This polymer gives internal quantum efficiencies of $\sim 2\%$ when used in place of polymer 3 in bilayer structures with PPV.

The demonstration of high efficiencies in polymer LEDs using metal electrodes which are not prone to oxidation represents an important step in the development of these devices. Long-term stability, both under storage and under drive, must be demonstrated before these devices can find applications, and there is much to be learned yet about the nature of the polymer-electrode interfaces. However, promising lifetime studies have been reported for indium tin oxide-PPV-Al devices, with operation without failure to 700 h²⁰, and recent studies of the Al-PPV interface indicate the formation of a stable covalently bonded structure²¹. □

Received 20 July; accepted 26 August 1993.

- Burroughes, J. H., Jones, C. A. & Friend, R. H. *Nature* **335**, 137-141 (1988).
- Burroughes, J. H. *et al. Nature* **347**, 539-541 (1990).
- Braun, D. & Heeger, A. J. *Appl. Phys. Lett.* **58**, 1982-1984 (1991).
- Ohmori, Y., Uchida, M., Muro, K. & Yoshino, K. *Jap. J. appl. Phys.* **30**, L1938-L1940 (1991).
- Ohmori, Y., Uchida, M., Muro, K. & Yoshino, K. *Jap. J. appl. Phys.* **30**, L1941-L1943 (1991).
- Grem, G., Leditzky, G., Ullrich, B. & Leising, G. *Adv. Mater.* **4**, 36-37 (1992).
- Burn, P. L. *et al. Nature* **356**, 47-49 (1992).
- Holmes, A. B. *et al. Synth. Met.* **55-57**, 4031-4040 (1993).
- Friend, R. H., Bradley, D. D. C. & Holmes, A. B. *Phys. World* **5**, 42-46 (1992).
- Tang, C. W. & VanSlyke, S. A. *Appl. Phys. Lett.* **51**, 913-915 (1987).
- Adachi, C., Tsutsui, T. & Saito, S. *Appl. Phys. Lett.* **57**, 531-533 (1990).
- Burn, P. L. *et al. in Electrical, Optical, and Magnetic Properties of Organic Solid State Materials* Vol. 245 (eds Chiang, L. Y., Garito, A. F. & Sandman, D. J.) 647-654 (Mater. Res. Soc. Symp. N, Pittsburgh, 1992).
- Brown, A. R. *et al. Appl. Phys. Lett.* **61**, 2793-2795 (1992).
- Ohmori, Y., Uchida, M., Muro, K. & Yoshino, K. *Solid St. Commun.* **80**, 605-608 (1991).
- Doi, S., Kuwabara, M., Noguchi, T. & Ohnishi, T. *Synth. Met.* **55-57**, 4174-4179 (1993).
- Lenz, R. W. & Handlovits, C. E. *J. org. Chem.* **25**, 813-817 (1960).
- Hörhold, H.-H. *Z. Chem.* **12**, 41-52 (1972).
- Burn, P. L. *et al. J. chem. Soc., Perkin Trans. 1*, 3225-3231 (1992).
- Helbig, M. & Hörhold, H.-H. *Makromol. Chem.* **194**, 1607-1618 (1993).
- Karg, S. *et al. in Proc. 6th Symp. on Unconventional Photoactive Solids* Leuven, Belgium, 1993; also as *Molec. Cryst. liq. Cryst.* (in the press).
- Dannetun, P. *et al. Synth. Met.* **55-57**, 212-217 (1993).

ACKNOWLEDGEMENTS. We thank the SERC, Cambridge Display Technology and the Nuffield Foundation (R.H.F.) for financial support. We acknowledge assistance from P. J. Hamer and D. R. Baigent, and from E. C. Constable and A. Cargill-Thompson.

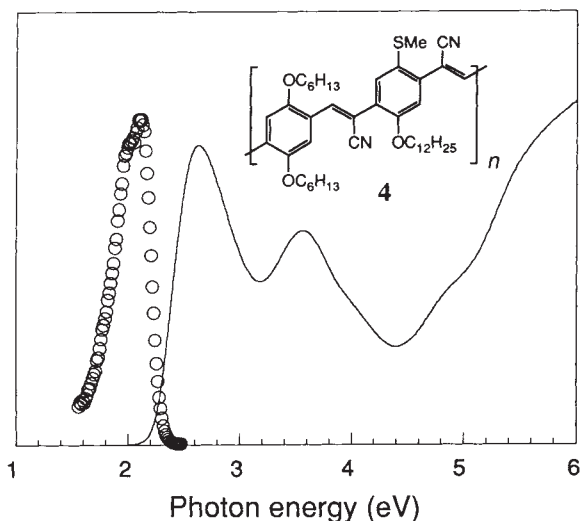


FIG. 4 Absorption (solid line) and emission (circles) spectra and structure of polymer 4, measured under the same conditions as for Fig. 2. The peak emission wavelength is 610 nm. (Vertical axis is in arbitrary units, linear scale.)

Prebiotic ammonia from reduction of nitrite by iron (II) on the early Earth

David P. Summers & Sherwood Chang

NASA-Ames Research Center, Moffett Field, California 94035, USA

THEORIES for the origin of life require the availability of reduced (or 'fixed') nitrogen-containing compounds, in particular ammonia. In reducing atmospheres, such compounds are readily formed by electrical discharges^{1,2}, but geochemical evidence suggests that the early Earth had a non-reducing atmosphere^{1,3-6}, in which discharges would have instead produced NO (refs 7-10). This would have been converted into nitric and nitrous acids and delivered to the early oceans as acid rain¹¹. It is known¹²⁻¹⁵, however, that Fe(II) was present in the early oceans at much higher concentrations than are found today, and thus the oxidation of Fe(II) to Fe(III) provides a possible means for reducing nitrites and nitrates to ammonia. Here we explore this possibility in a series of experiments which mimic a broad range of prebiotic seawater conditions (the actual conditions on the early Earth remain poorly constrained). We find

that the reduction by Fe(II) of nitrites and nitrates to ammonia could have been a significant source of reduced nitrogen on the early Earth, provided that the ocean pH exceeded 7.3 and is favoured for temperatures greater than about 25 °C.

To determine the reactivity of nitrite (NO₂⁻) with aqueous Fe(II) a solution of both was allowed to react (equation (1)) and the formation of ammonia was monitored over time. All reaction rates



presented here are initial rates. The total amount of ammonia produced, when the reaction was complete, was used to calculate the percentage of nitrite converted to ammonia (the product yield). All ammonia concentrations are relative to samples taken at the start of the experiment. No ammonia is detected in the absence of nitrite, nitrate (NO₃⁻) or Fe(II).

The data in Table 1 confirm earlier observations¹⁶ that Fe(II) will reduce nitrite and nitrate to ammonia. During the reaction, which proceeds in the absence of light, the clear pale yellow-green solution becomes an optically dense dark-green suspension of ferrous hydroxide ('green rust'). The measured rates are faster than those implied in earlier work¹⁶. In addition to our results on nitrite reduction reported here, some reduction of nitrate to ammonia was detected. Ammonia was not, however, formed reproducibly from nitrate in every experiment, although product yields were ~95% when the reaction occurred. In all cases, ammonia formation from nitrate was slower, by at least a factor of eight, than from nitrite.

Because conditions that might affect the reaction on the prebiotic earth are not known, the kinetics were studied as a function of temperature, pH and concentrations of Fe(II) and nitrite. Data in Table 1 show the dependence of the rate of reaction on pH (entries 1–8). At pH ≤ 7.3, the reaction does not proceed at a detectable rate. The maximum rate is found at a pH of 7.6. As the pH increases above 7.6 the rate falls off, making reactivity above ~pH 9 low. The rate also follows Arrhenius behaviour with respect to temperature (at least between 0 and 40 °C), yielding a slope of 22,000 K⁻¹. Data in Table 2 show that the rate varies in proportion to the concentration of nitrite but in a more complex manner with the concentration of Fe(II). A least-squares fit of the equation $y = kx^n$ to the data on rate as a function of [Fe²⁺] yields a value of 1.8 for n , thus giving an apparent rate law (at pH 7.9) of the form $\text{rate} = k[\text{NO}_2^-][\text{Fe}^{2+}]^{1.8}$, where $k = 35 \times 10^{-6} \text{ min}^{-1} \text{ M}^{-1.8}$.

The experiments described were conducted in solutions that were made from pure water. The ocean today contains large

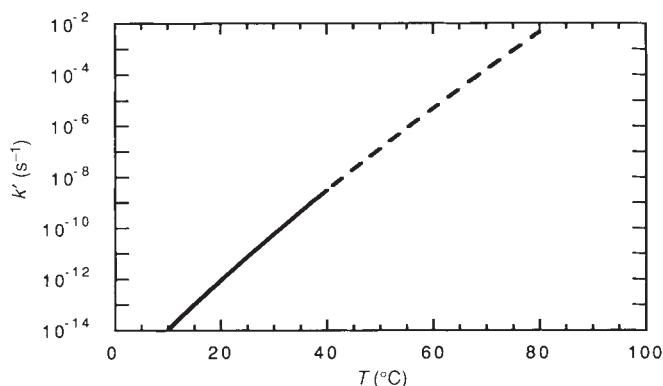


FIG. 1 Plot of the pseudo-first-order rate constant ($k' = k[\text{Fe(II)}]^{1.8}$) as a function of ocean temperature (pH 7.6, 0.2 atm CO₂, saturated with respect to siderite). The rate constant k was measured experimentally between 0 and 40 °C (solid line) and extrapolated to 80 °C (dashed line); see text. Values for the K_{sp} (solubility product) of Fe(CO₃) are from ref. 22.

TABLE 1 Rate of reduction of nitrite to ammonia by aqueous Fe(II)*

Entry	T (°C)	pH	Additions†	Rate (μM min ⁻¹)	Prod. yield‡(%)
1	22	7.1	None	<0.004§	—
2	22	7.3	None	<0.008§	—
3	22	7.5	None	10	35
4	22	7.6	None	15	37
5	22	7.9	None	8.8	28
6	22	8.0	None	8.5	33
7	22	8.0	NaCl	44	64
8	22	8.4	NaCl	6.6	25
9	22	8.0	Cations	67	85
10	22	8.0	Anions	36	52
11	0	7.6	None	~0.1	3.8
12	11	7.6	None	0.45	24
13	23	7.6	None	12	44
14	30	7.6	None	61	74
15	40	7.6	None	630	60
16	22	7.5	CO ₂	0	0
17	80	7.5	CO ₂	0.006	75

* A volume of a nitrogen-purged NaNO₂ solution (23 mM) sufficient to make up a 0.32 mM solution was added by syringe to a 12.3 mM FeCl₂ (Aldrich) solution (350 ml) stirred under nitrogen purging. In occasional experiments a volume of FeCl₂ solution sufficient to make up 12.3 mM was added to a 0.32 mM solution of NaNO₂ with no change in results. After a ($t=0$) blank was withdrawn and purging stopped (to avoid sweeping out ammonia), the reaction was allowed to proceed under nitrogen. Aliquots were withdrawn by syringe and analysed for ammonia by colorimetric methods¹⁴.

† Concentrations: NaCl, 0.53 M NaCl; anions, 5 mM NaBr, 28 mM Na₂SO₄ and 0.53 M NaCl; cations, 51 mM MgCl₂, 12 mM CaCl₂, 13 mM KCl and 0.53 M NaCl; CO₂, saturated with respect to FeCO₃, 1 atm CO₂, and 0.53 M NaCl.

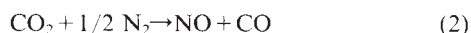
‡ The percentage of reacted nitrite that is converted to ammonia.

§ These values are within experimental noise and represent upper limits. All other numbers and changes from experiment to experiment are significant.

amounts of sodium chloride as well as small amounts of other ions, including K⁺, Mg²⁺, Ca²⁺, Br⁻ and SO₄²⁻. The presence of these salts in the early ocean may have affected the kinetics of nitrite reduction. Anions may interfere with the reaction, particularly by binding to the Fe(II) centres and deactivating them. Cations may complex nitrite or anionic iron complexes. Increased pressure of carbon dioxide would have led to increased bicarbonate concentrations, and the concentration of sulphate is likely to have been lower than it is today¹³. The concentration of halide ions in the early ocean is not well known, but indications are that it was not far from the current value¹².

When 0.53 M NaCl is added to the reaction, a fivefold increase in the rate is observed (entries 6 and 7, Table 1). When other cations (that exist in sea water now) are added, an additional rate increase is observed (entries 7 and 9). The presence of sulphate and bromide ions at levels equal to those found in present-day environments reduced the rate of reduction slightly (entries 7 and 10). The situation with bicarbonate is more complicated. The carbon dioxide pressure (and carbonate/bicarbonate concentration) was likely to have been much higher in the early atmosphere than it is today, with suggested values ranging from 0.2–10 atm^{3,6,17}. Reactions in this range are difficult to run because the limited solubility of siderite (FeCO₃) required very low concentrations of Fe(II). Two experiments were conducted in solutions that had been saturated with sodium bicarbonate, one at 25 °C and one at 80 °C (entries 16 and 17). At the end of two months the room-temperature experiment showed no production of ammonia. In the 80 °C experiment, however, nitrite was reduced at a rate of 0.006 μM min⁻¹. If we correct for temperature (using the Arrhenius equation) and the concentration of Fe(II), this rate is what would be expected, within experimental error, from the rate law given previously.

With these data we can calculate a possible rate of ammonia formation in the early ocean. Shock heating produces NO (ref. 7), for instance by equation (2):



We used a production rate for NO of $1.1 \times 10^{11} \text{ mol yr}^{-1}$ from shock heating by meteors passing through the atmosphere 4.0 Gyr ago (see ref. 9 and refs therein). To calculate the production rate of NO from lightning, we have taken the most recent estimate of lightning frequency¹⁸, together with an estimate of the efficiency with which NO is produced from lightning discharges⁷, to obtain $1.4 \times 10^{11} \text{ mol yr}^{-1}$. Next, we assume that all the NO is washed into the ocean¹¹. Depending on the stoichiometry of the reactions involved, 20–50% of the NO is converted to nitrite^{11,19,20}. Choosing an intermediate value of 33%, we obtain a nitrite production rate of $4.9 \times 10^{10} \text{ mol yr}^{-1}$. We will consider an ocean at 25 °C, under 0.2 atm of carbon dioxide, at pH 7.6. This would limit the concentration of Fe(II) to 190 nM, based on siderite solubility from equilibrium constants²¹ and thermodynamic data for FeCO₃ (ref. 22). Under these conditions, and assuming a steady state between atmospheric production of nitrite and its reduction to ammonia, the rate of reduction is fast enough to hold the concentration of nitrite to 1.6 μM.

The reaction rate shows a strong dependence on temperature. At higher temperature the solubility of siderite increases²² ([Fe²⁺] should also depend inversely on [CO₃²⁻]). If the concentration of Fe(II) is fixed by the solubility of FeCO₃, then we can define a pseudo-first-order rate law, $\text{rate} = k'[\text{NO}_2^-]$, where $k' = k[\text{Fe(II)}]$ ¹⁸. A plot of k' against temperature (at a partial pressure of CO₂ of 0.2 atm.) is shown in Fig. 1. For this plot, we have artificially held the pH at 7.6 to separate the effects of temperature and pH on k' . Between 10 and 80 °C, the rate of nitrite reduction in the ocean increases by 11 orders of magnitude. Clearly, reduction is strongly favoured by warm environments. For example at 80 °C, the steady-state concentration of nitrate drops to $4.1 \times 10^{-15} \text{ M}$.

One possible sink for nitrite, competing with reduction to ammonia, is destruction in waters passing through oceanic hydrothermal systems. At present, the timescale for ocean cycling through these systems is ~10 million years²³, but we adopt a shorter duration to take into account increased tectonic and/or volcanic activity early on^{24,25}. The fate of nitrite in hydrothermal systems is not known. If we assume complete destruction to nitrogen as a worst-case model, and if the oceans cycled through the hydrothermal systems every 2.5 million years, the maximum hydrothermal sink for nitrite would be $\sim 3 \times 10^8 \text{ mol yr}^{-1}$. This hydrothermal sink is a factor of 150 smaller than nitrite reduction in the bulk ocean at an ocean temperature of 25 °C; at an ocean temperature of 80 °C, it is a factor of 3×10^{10} lower.

Another competing abiotic sink for nitrite is aqueous photochemical destruction in near-surface ocean waters. It is difficult to assess the importance of this sink. The amount of near-ultraviolet light reaching the ocean surface in the prebiotic environment is uncertain²⁶. One product of photolysis is NO, which can re-enter the atmosphere and be reconverted to nitrite/nitrate. Also, in pure water, there is no net reaction because the primary photoproducts (NO and OH⁻) back-react to reform nitrite with 100% efficiency²⁷. Net photochemical destruction in the ocean is dependent on the presence of poorly defined trace species²⁷, which were almost certainly present at different concentrations (or missing completely) in the early Archaean ocean. We can use photochemical rates occurring in the contemporary ocean²⁷ to provide an upper limit to the size of the photochemical sink. This limit is 1,200 times higher than nitrite reduction at room temperature, but a factor of 1.2×10^5 lower at 80 °C. Again, warmer temperatures strongly favour reduction to ammonia. Assuming a modern oceanic vertical mixing rate of 3 m yr^{-1} (ref. 28), then ammonia, iron and nitrite concentrations would have been fairly uniform over the whole ocean.

The steady-state concentration of ammonia (present mostly

TABLE 2 Dependence of the rate of nitrite reduction by aqueous Fe²⁺ on nitrite and Fe(II) concentration*

[NO ₂ ⁻] (mM)	[Fe ²⁺] (mM)	Rate (μM min ⁻¹)	pH	Relative concentration	Relative rate	Product yield† (%)
Nitrate concentration dependence						
0.0327	12.3	0.9	7.9	1	1	—
0.131	12.3	5.0	7.9	4	5.6	34
0.327	12.3	8.8	7.9	10	9.8	28
0.654	12.3	18	7.9	20	20.0	30
0.327	12.3	4.2	8.0	1	1	28
0.654	12.3	7.8	8.0	2	1.9	26
Fe(II) concentration dependence						
0.327	3.2	1.3	7.9	1	1	—
0.327	4.9	1.3	7.9	1.5	1	27
0.327	8.9	7.7	7.9	2.8	6.1	30
0.327	12.3	8.8	7.9	4	6.9	28
0.327	19.7	27	7.9	6.2	21	42

* Values are measured room temperature. Variable volumes of nitrogen-purged 23 mM NaNO₂ (Baker) were added by syringe to an FeCl₂ 12.3 mM; (Aldrich) solution (350 ml) stirred under nitrogen purging. After a (t=0) blank was withdrawn and purging stopped (to avoid sweeping out ammonia), the reaction was allowed to proceed under nitrogen. Aliquots were withdrawn by syringe and analysed for ammonia by colorimetric methods¹⁴.

† The percentage of reacted nitrite that is converted to ammonia.

as ammonium, NH₄⁺) would depend on what sinks for ammonia were active. Again we assume an ocean temperature of 25 °C and a pH of 7.6. If the predominant sink for ammonia was hydrothermal decomposition (such as was outlined for nitrite decomposition above) then, with a production rate of $3.6 \times 10^{10} \text{ mol yr}^{-1}$, the steady-state ammonia concentration would have been 70 μM.

An alternative sink is gas-phase photochemical destruction of ammonia in the atmosphere, as modelled by Kasting²⁹. We assume ammonia in the ocean to be in equilibrium with ammonia in the atmosphere. In this case, the steady-state concentration of ammonium ions in the ocean would have been 3.6 μM. If other sinks were active, this value would be lower. Although sequestration in clays represents a possible sink, there is reason to believe that this process would only have set an upper limit of 0.01 M on the NH₄⁺ concentration³⁰.

The reduction of nitrite to ammonia represents a plausible process if oceanic pHs were greater than 7.3. The reaction is also strongly favoured by warmer oceans^{17,31,32}. Whether ammonia concentrations of the order of 3.6–70 μM in the ocean would have been adequate for prebiotic evolution remains uncertain². □

Received 10 May; accepted 24 August 1993.

- Schoff, J. W. (ed.) *Earth's Earliest Biosphere* 53–92 (Princeton Univ. Press, 1983).
- Stribling, R. & Miller, S. L. *Orig. Life Evol. Biosph.* **17**, 261–273 (1978).
- Walker, J. C. G. *Orig. Life Evol. Biosph.* **16**, 117–127 (1985).
- Mattoli, G. S. & Wood, B. J. *Nature* **322**, 626–628 (1986).
- Gregor, C. B., Garrels, R. M., Mackenzie, F. T. & Maynard, J. B. (eds) *Chemical Cycles in the Evolution of the Earth* 42–79 (Wiley, New York, 1988).
- Kasting, J. F. *Precamb. Res.* **34**, 205–229 (1987).
- Chameides, W. L. & Walker, J. C. G. *Orig. Life Evol. Biosph.* **11**, 291–302 (1981).
- Yung, Y. L. & McElroy, M. B. *Science* **203**, 1002–1004 (1979).
- Kasting, J. F. *Orig. Life Evol. Biosph.* **20**, 199–231 (1990).
- Fegley, B. Jr, Prinn, R. G., Hartman, H. & Watkins, G. H. *Nature* **319**, 305–308 (1986).
- Mancinelli, R. L. & McKay, C. P. *Orig. Life Evol. Biosph.* **18**, 311–325 (1988).
- H. D. Holland *Econ. Geol.* **68**, 1169–1172 (1973).
- Walker, J. C. G. & Brimblecombe, P. *Precamb. Res.* **28**, 205–222 (1985).
- Derry, L. A. & Jacobsen, S. B. *Geochim. Cosmochim. Acta* **54**, 2965–2975 (1990).
- Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* Ch. 4 (Princeton Univ. Press, 1989).
- Buresh, R. J. & Moraghan, J. T. *J. envir. Qual.* **5**, 320–324 (1976).
- Walker, C. G. J. et al. in *Earth's Earliest Biosphere* (ed. Schoff, J. W.) 260–284 (Princeton Univ. Press, 1983).
- Chyba, C. & Sagan, C. *Orig. Life Evol. Biosph.* **21**, 3–17 (1991).
- Grätzel, Von M., Taniguchi, S. & Henglein, A. *Ber. Bunsenges. phys. Chem.* **74**, 1003–1010 (1970).
- Doyle, M. P. & Mahapatro, S. N. *J. Am. chem. Soc.* **106**, 3678–3679 (1984).
- Butler, J. N. *Carbon Dioxide Equilibria and Their Applications* (Addison-Wesley, Reading, MA, 1982).
- Freier, R. K. (ed.) *Aqueous Solutions Vol 1* (de Gruyter, Berlin, 1976).

23. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* 205 (Princeton Univ. Press).
24. Windley, B. F. (ed.) *The Early History of the Earth* (Wiley, New York, 1976).
25. Turcotte, D. L. *Earth planet. Sci. Lett.* **48**, 50–53 (1980).
26. Kasting, J. F., Zahnle, K. J., Pinto, J. P. & Young, A. T. *Orig. Life Evol. Biosph.* **19**, 95–108 (1989).
27. Zafiriou, O. C. & True, M. B. *Mar. Chem.* **8**, 9–32 (1979).
28. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* 243 (Lamont–Doherty geol. Obs. Palisades, 1982).
29. Kasting, J. F. *J. geophys. Res.* **87**, 3091–3098 (1982).
30. Bada, J. L. & Miller, S. L. *Science* **159**, 423–425 (1968).
31. Kuhn, W. E. & Kasting, J. F. *Nature* **301**, 53–55 (1983).
32. Walker, C. G. J. *Orig. Life Evol. Biosph.* **16**, 117–127 (1985).

ACKNOWLEDGEMENTS. We thank J. Kasting and D. Des Marais for discussions. D.P.S. was supported by a senior research fellowship from the National Research Council. Partial support for this work was provided by NASA.

Upper-crustal strength inferred from stress measurements to 6 km depth in the KTB borehole

Mark D. Zoback*, R. Apel†, J. Baumgärtner‡, M. Brudy†, R. Emmermann§, B. Engeser§, K. Fuchs†, W. Kessels§, H. Rischmüller§, F. Rummel|| & L. Vernik*

* Department of Geophysics, Stanford University, Stanford, California 94305, USA

† Geophysical Institute, University of Karlsruhe, 76187 Karlsruhe, Germany

‡ SOCOMINE, F-67250 Soultz sous Forêts, France

§ KTB Project, Geological Survey of Lower Saxony, 30655 Hannover, Germany

|| Department of Geophysics, Ruhr-University Bochum, 44780 Bochum, Germany

It has been suggested^{1–6} that in many cases the average strength of the continental crust is quite low (tens of megapascals), so that the crust has little effect on the large-scale deformation of the lithosphere. But laboratory friction studies^{7,8}, combined with simple faulting theory^{9,10} (as well as extrapolation of *in situ* stress measurements from the upper 3 km of the crust¹¹), imply that if pore pressure is approximately hydrostatic at mid-crustal depth, crustal strength is appreciable (hundreds of megapascals) and would markedly constrain the nature of lithospheric deformation^{12–15}. Here we report estimates of the magnitude of *in situ* stresses to 6 km depth in the KTB borehole in southern Germany. Our results indicate a high-strength upper crust, in which the state of stress is in equilibrium with its frictional strength. We suggest that plate-driving forces in the continental lithosphere in this part of western Europe are transmitted principally through the upper crust, and that this may also be the case in other continental areas of moderate to elevated heat flow.

The KTB (Kontinentales Tiefbohrprogramm der Bundesrepublik Deutschland) drillsite is located in an area of occasional minor seismicity where highly deformed Palaeozoic metamorphic rocks have been intruded by Variscan (late Palaeozoic) granites. In the first phase of the drilling project (completed in 1989), a pilot hole was continuously cored to a depth of 4 km. It encountered a highly folded and steeply dipping sequence of foliated gneisses and intercalated massive amphibolites. Stress measurements in the pilot hole^{16,17} show a NNW-oriented direction of maximum horizontal principal stress and strike-slip to normal faulting stress magnitudes. These quantities are similar in orientation and relative magnitude to those found in much of central Europe^{18,19}. The second phase of the project involves

drilling a much larger diameter borehole to a total depth of 10 km. Below 3.2 km in the main borehole, nearly all the rock is relatively massive amphibolite from apparently the same rock units encountered at shallow depth.

An integrated stress-measurement strategy was developed to determine *in situ* stress magnitudes to 10 km depth in the main borehole. An important part of this strategy was to propagate a hydraulic fracture away from a ~20-m-long section of open hole that was drilled after steel casing was cemented into the hole at 6 km. From such a 'hydrofrac' test one can determine the magnitude of the least principal stress from the so-called 'shut-in' pressure²⁰ and low-flow-rate pumping pressure²¹. We assume that the least principle stress corresponds to S_{hmin} , the least horizontal principal stress, as one principal stress generally appears to be vertical in the upper crust (corresponding to the lithostat, S_v ²²) and in the KTB boreholes there are ubiquitous near-vertical drilling-induced tensile fractures (described below) which indicate a near-vertical principal stress²³. One cannot obtain confident estimates of the maximum horizontal principal stress, S_{Hmax} , from the type of hydrofrac test conducted at 6 km because fracture initiation is likely to be affected by the irregular shape of the borehole wall caused by stress-induced wellbore breakouts (described below) and/or pre-existing fractures. The pressure record from which the magnitude of S_{hmin} was determined is shown in Fig. 1a. Both the shut-in pressure and low-flow-rate pumping pressure indicate that the least principal stress has a magnitude of 114 ± 5 MPa.

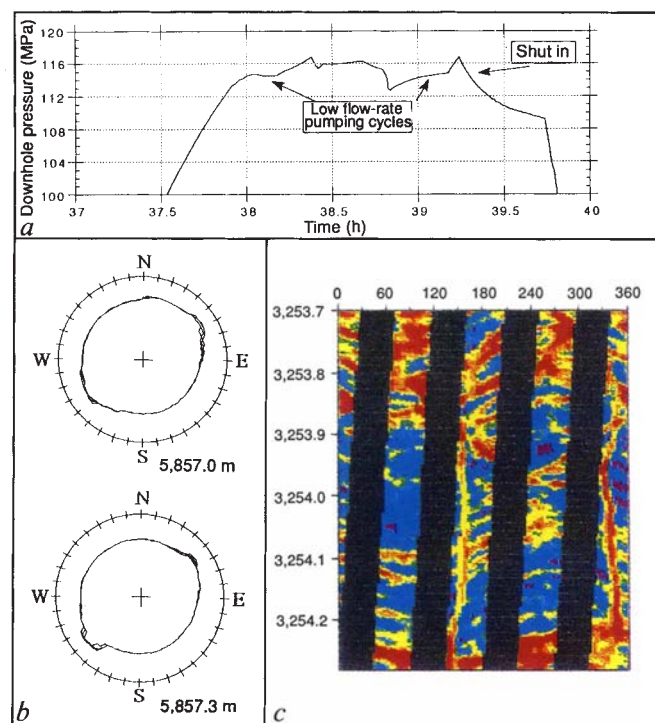


FIG. 1 a, Downhole pressure recorded during the hydraulic fracturing test. b, Cross-sectional shape of the borehole in an interval of stress-induced wellbore breakouts. Note that the breakouts occur ~180° apart (as expected from the concentration of elastic stress around the wellbore). The azimuth of elongation corresponds to the direction of least horizontal compression. c, A fine-scale electrical conductivity image of the inside of the borehole wall. Vertical axis is depth in metres, horizontal axis is azimuth around the wellbore with respect to north, colour indicates fine-scale electrical conductivity (red is high conductivity, yellow is intermediate, purple is low) and black bands indicate sections of the borehole wall where no data was obtained. Near-vertical drilling-induced tensile fractures cut across foliation planes in the rock. As theoretically expected, the tensile fractures occur 180° apart, in the direction of maximum horizontal compression.

To estimate the magnitude of S_{Hmax} , the integrated stress-measurement strategy uses the hydrofrac-determined value of S_{hmin} coupled with analysis of compressive and tensile failures around the wellbore. These failures result from the stress concentration associated with drilling an initially circular hole into a pre-stressed medium²⁴. Detailed inspection of the KTB pilot and main boreholes with ultrasonic and electrical imaging devices indicates that there are many depth intervals where the rock surrounding the borehole is simultaneously failing in compression and tension as it is being drilled. That is, stress-induced compressive failures (or wellbore breakouts) (Fig. 1b) are observed at the azimuth of S_{hmin} (where the stress concentration is most compressive) and drilling-induced tensile fractures (Fig. 1c) are observed at the azimuth of S_{Hmax} (90° from the breakouts) where the stress concentration is least compressive.

To estimate S_{Hmax} from the occurrence of wellbore breakouts, we used information regarding (1) the least principal stress determined from hydraulic fracturing, (2) rock strength determined from laboratory measurements on cores, and (3) the circumferential angle over which the wellbore failed, as determined from detailed analysis of borehole televiewer data. The theoretical basis for this technique and its detailed application to the KTB pilot hole is described elsewhere^{25,26}. Uncertainties in the computed values of S_{Hmax} (Fig. 2) result from the combined uncertainties in rock strength and variations of the shape of the breakouts around the wellbore.

Drilling-induced tensile failures at the borehole wall arise from the combined effects of the stress concentration around the borehole, cooling-induced tensional stresses (resulting from circulation of relatively cold drilling mud into the hole) and the excess pressure in the borehole during drilling²⁷. The near-vertical tensile fractures in the KTB pilot and main boreholes are clearly drilling-induced. They are not present in the cores, but are observed (as theoretically expected) at the azimuth of S_{Hmax} and cut across pre-existing planes of weakness such as foliation planes. Determination of S_{Hmax} from the drilling-induced tensile fractures involved several steps. First, we assumed that the rock

has no tensile strength because these fractures could initiate at a pre-existing flaw around the wellbore. Second, because it was not known how much cooling was necessary to induce the observed thermal fractures, we calculated maximum and minimum values of S_{Hmax} corresponding to the full range of temperatures and excess borehole pressures during drilling at which the fracture might have initiated. The upper bound for S_{Hmax} is calculated by assuming that the fractures occurred with no borehole cooling (that is, simply as a result of the tectonic stress concentration) and no excess fluid pressure during drilling, whereas the lower bound is calculated by assuming that the fractures occurred after the maximum amount of borehole cooling occurred (Fig. 2). (More details of the analysis of drilling-induced fractures are contained in a preprint by R.A., M.D.Z. and K.F., see Supplementary Information.)

Because the magnitude of S_{hmin} and S_{Hmax} are known to 3 km depths from conventional hydraulic fracturing stress measurements in the KTB pilot hole (ref. 16; also available as Supplementary Information), we can compare the values of S_{Hmax} obtained by analysis of the breakouts and tensile fractures with that determined from hydraulic fracturing, and thus effectively calibrate the integrated stress-measurement strategy. As shown in Fig. 2, the S_{Hmax} values determined from the breakout analysis compare extremely well with the direct hydrofrac-determined values in the pilot hole²⁸. A similarly good match between breakout-determined and hydrofrac-determined S_{Hmax} values was found in the 3.5-km-deep Cajon Pass scientific research borehole²⁵. The data required to estimate S_{Hmax} from the tensile fractures in the KTB pilot hole were available only at depths between 3 and 4 km. The values of S_{Hmax} determined from the tensile fractures in that interval compare extremely well with the values of S_{Hmax} estimated both from the breakout analysis at the same depths and from extrapolation of the hydrofrac-determined S_{Hmax} values.

In the interval between 4 and 6 km, we used an analysis of both the drilling-induced tensile fractures and breakouts to obtain the range of values for S_{Hmax} shown in Fig. 2. Although the uncertainty in S_{Hmax} at great depth is appreciable (reflecting the cumulative uncertainties in the various parameters noted above), a strike-slip faulting stress regime is clearly indicated (the vertical stress is the intermediate principal stress) and the maximum differential stress ($S_{Hmax} - S_{hmin}$) is substantial (>100 MPa).

The maximum differential stresses determined in the KTB pilot and main boreholes are summarized in Fig. 3a and compared to theoretical estimates of the maximum principal stress difference in the crust. The strength profile in Fig. 3a is for the case of strike-slip faulting in the upper crust (appropriate for the KTB region²⁹ and consistent with the *in situ* stress measurements), a coefficient of friction (μ) of 0.7 (ref. 7) and hydrostatic pore pressure (which is consistent with pore pressure measurement in the KTB borehole to 6 km depth). To derive the linear increase of frictional strength with depth in the upper crust, we calculated the value of S_{Hmax} required to cause frictional sliding on pre-existing fault planes for the measured values of S_{hmin} . Based on these assumptions, the maximum differential stress reaches values of ~300–400 MPa in the mid-crust. As shown in Fig. 3a, the differential stress values determined to 6 km in the KTB main borehole are consistent with the strength of the upper crust based on laboratory measurements and simple frictional faulting theory.

To consider the strength of the upper crust in the context of how the lithosphere transmits plate-driving forces, it is necessary to estimate the ductile creep strength of the lower crust and upper mantle. To do this we must estimate temperature at depth, assume an appropriate strain rate and use appropriate rheological values determined from laboratory experiments. Temperatures in the lower crust and upper mantle are based on a conductive thermal model based on the observed geotherm and relatively high heat flow (~74 mW m⁻²) observed at the KTB

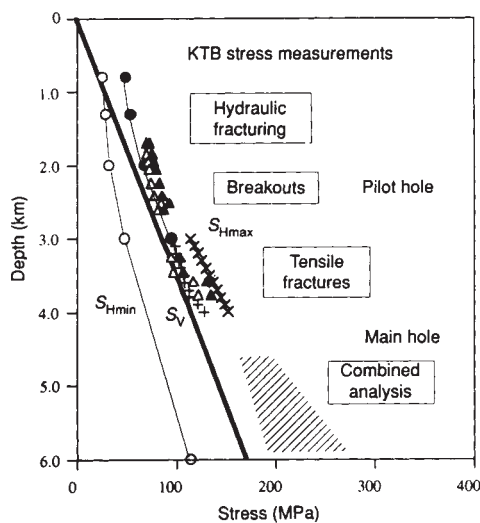


FIG. 2 S_{hmin} and S_{Hmax} values in the KTB pilot hole and main borehole. The values determined from the hydraulic fracturing measurements are indicated by $\circ$ (S_{hmin}) and $\bullet$ (S_{Hmax}). The upper- ($\blacktriangle$) and lower-bound ($\triangle$) estimates of the magnitude of S_{Hmax} determined from wellbore breakouts (as well as the hydrofrac-determined S_{hmin} values) are also shown. The upper and lower bound S_{Hmax} values determined from analysis of drilling-induced tensile fractures are indicated by + and $\times$, respectively. The combined analysis (shown by the hatched area), refers to the upper and lower bound estimates of S_{Hmax} between 4.5 and 6 km depth in the main borehole which are consistent with the occurrence of both tensile fractures and breakouts. The vertical stress is based on the average measured density of 2.8 g cm⁻³.

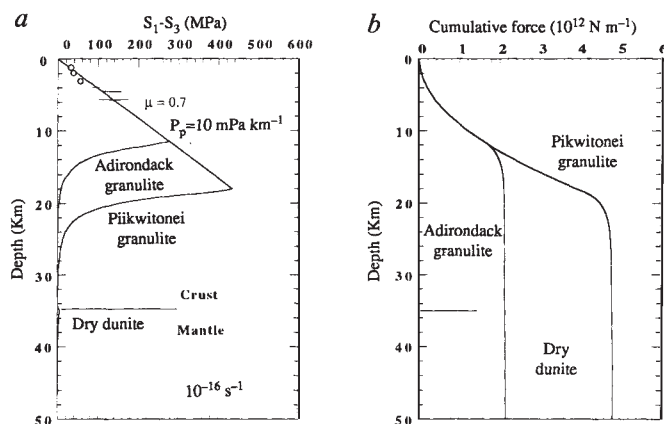


FIG. 3 a, Theoretical maximum differential stress ($S_1 - S_3$) profiles for conditions similar to those in the region of the KTB drillsite. Data points shown as $\circ$ indicate the difference between S_{Hmax} and S_{Hmin} as determined from hydraulic fracturing data to 3 km depth in the pilot hole. The data points shown as horizontal lines indicate a generalization of the values shown in Fig. 2 from the breakout and tensile fracture analysis. A strike-slip faulting stress regime with hydrostatic pore pressure and a coefficient of friction (μ) of 0.7 was used to calculate the theoretical strength in the upper crust (see text). b, Cumulative crustal force corresponding to the strength profiles shown in a).

site and in this part of central Europe. The values of thermal conductivity, radioactive heat production and the rate at which heat production decays are assumed to be typical values for continental crust^{30,31}. We use the creep properties of Adirondack and Pikwitonei granulites as indicative of lower crustal rocks with very low and very high ductile strengths, respectively, at equivalent temperatures and strain rates³². Although the depth of the brittle-ductile transition implied by the lower-strength Adirondack granulite is reasonably consistent with the maximum depth (~ 12 km) of earthquakes in the region, it is unwise to place too much significance on this correlation as mechanisms other than the onset of crystal plasticity may play an important role in determining the depth of the brittle-ductile transition³³. To obtain a maximum value of the ductile strength of the lower crust and upper mantle, we assume an average strain rate of 10^{-16} s^{-1} , which is a reasonable upper limit for plate interiors (at higher strain rates, plates would deform too much for plate reconstruction to be possible) and that the strength of the upper mantle corresponds to that of 'dry' dunite³². Note that despite these assumptions, the upper mantle deforms at extremely low differential stress levels because of its high temperatures.

The cumulative force required to cause lithospheric deformation (Fig. 3b) is equivalent to the area under the strength curves shown in Fig. 3a (ref. 3). Note that it takes a cumulative force of $2.5 \times 10^{12} \text{ N m}^{-1}$ to cause steady-state deformation of the lithosphere and that cumulative lithospheric strength is largely due to that of the upper, brittle crust.

The primary source of tectonic stress in western Europe appears to be related to the ridge-push force^{18,34} resulting from the cooling and thickening of the lithosphere away from the Mid-Atlantic Ridge. The magnitude of ridge push averaged over the thickness of the lithosphere is $\sim 2.5 \times 10^{12} \text{ N m}^{-1}$ (refs 35, 36). Because the available tectonic force is comparable in magnitude to the cumulative lithospheric strength, widespread seismicity in this part of western Europe appears to reflect steady-state failure equilibrium (at very low strain rate); force transmission through the lithosphere is largely in the upper 10–20 km of the crust, effectively in an upper-crustal stress guide. A similar process may also occur in the seismically active part of eastern North America as ridge-push appears to be the primary source of tectonic stress there^{34,37} and heat flow is moderately high³¹.

One can extend these arguments to geologically stable shield areas and relatively old ocean basins adjacent to western Europe and eastern North America, which are subjected to essentially the same lithospheric forces. As these regions are characterized by very low heat flow ($\sim 40 \text{ mW m}^{-2}$), upper-mantle strength is expected to be quite high (because of relatively low temperatures) and capable of supporting much of the total force acting on the lithosphere. Hence, significant lithospheric deformation would not be expected to occur because the cumulative strength of the lithosphere would be appreciably greater than the force available to cause deformation. $\square$

Received 25 Jan; accepted 23 August 1993.

- England, P. C. & McKenzie, D. P. *Earth planet. Sci. Lett.* **70**, 295–321 (1982).
- McKenzie, D. P. & Jackson, J. A. *Earth planet. Sci. Lett.* **65**, 182–202 (1983).
- Houseman, G. A. & England, P. C. *J. geophys. Res.* **91**, 3651–3663 (1986).
- Jackson, J. A. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **93**, 45–73 (1988).
- Sonder, L. & England, P. *Earth planet. Sci. Lett.* **77**, 81–90 (1986).
- Molnar, P. in *Fault Mechanics and Transport Properties of Rock* (eds Evans, B. & Wong, T.-F.) (Academic, London, 1992).
- Byerlee, J. D. *Pure appl. Geophys.* **116**, 615–629 (1978).
- Kirby, S. J. *geophys. Res.* **85**, 6353–6363 (1980).
- Sibson, R. H. *Nature* **249**, 542–544 (1974).
- Brace, W. F. & Kohlstedt, D. L. *J. geophys. Res.* **85**, 6248–6252 (1980).
- Zoback, M. D. & Healy, J. H. *J. geophys. Res.* **97**, 5039–5057 (1992).
- Goetze, C. & Evans, B. *Geophys. J. R. astr. Soc.* **59**, 463–478 (1979).
- McNutt, M. J. *geophys. Res.* **89**, 11180–11194 (1984).
- Kuznir, N. J. & Park, R. G. *Geophys. J. R. astr. Soc.* **79**, 513–538 (1984).
- Cloetingh, S. & Wortel, R. *Tectonophysics* **132**, 49–67 (1986).
- Baumgärtner, J., Rummel, F. & Zoback, M. D. in *KTB Report No. 90-6a* (Kontinentales Tiefbohrprogramm, Bundesrepublik Deutschland, Hannover, 1990).
- Mastin, L. G., Heinemann, B., Krammer, A., Fuchs, K. & Zoback, M. D. *Scient. Drilling* **2**, 1–12 (1991).
- Müller, B., et al. *J. geophys. Res.* **97**, 11783–11804 (1992).
- Zoback, M. L. *J. geophys. Res.* **97**, 11703–11728 (1992).
- Haimson, B. & Fairhurst, C. in *Situ Stress Determination at Great Depth by Means of Hydraulic Fracturing* (Soc. Min. Engrs of Am. Inst. Mech. Engrs, New York, 1970).
- Hickman, S. H. & Zoback, M. D. in *Hydraulic Fracturing Stress Measurements* (eds Zoback, M. D. & Haimson, B. C.) 44–54 (Nat. Acad., Washington DC, 1983).

- Zoback, M. L. & Zoback, M. D. in *Geophysical Framework of the Continental United States* (ed. Pakiser, L. A. M. W.) (Geol. Soc. Am., Boulder, 1989).
- Brudy, M. & Zoback, M. D. *Int. J. Rock Mech. Mining Sci.* (in the press).
- Kirsch, G. Z. *Ver. Dt. Ing.* **42**, 797–807 (1989).
- Vernik, L. & Zoback, M. D. *J. geophys. Res.* **97**, 5109–5119 (1992).
- Vernik, L., Lockner, D. & Zoback, M. D. *Scient. Drilling* **3**, 153–160 (1992).
- Moos, D. & Zoback, M. D. *J. geophys. Res.* **95**, 9305–9325 (1990).
- Vernik, L., Zoback, M. D. & Brudy, M. *Scient. Drilling* **3**, 161–169 (1992).
- Grünthal, G. & Stromeyer, D. *Gerlands Beiträge zur Geophysik* **95**, 443–452 (1986).
- Smithson, S. B. & Decker, E. R. *Earth planet. Sci. Lett.* **22**, 215–225 (1974).
- Morgan, P. & Gosnold, W. D. in *Geophysical Framework of the Continental United States* (Geol. Soc. Am. Memoir No. 72) (eds Pakiser, L. C. & Mooney, W. D.) 493–522 (Geol. Soc. Am., Boulder, 1989).
- Carter, N. L. & Tsenn, M. C. *Tectonophysics* **136**, 27–63 (1987).
- Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. *Geophys. Res. Lett.* **18**, 609–612 (1991).
- Zoback, M. L. et al. *Nature* **341**, 291–298 (1989).
- Dahlen, F. A. *J. geophys. Res.* **86**, 7801–7807 (1981).
- Fleitout, L. & Froidevaux, C. *Tectonics* **2**, 315–324 (1983).
- Zoback, M. D. & Zoback, M. L. in *Neotectonics of North America* (eds Slemmons, D. B., Engdahl, E. R., Zoback, M. D. & Balckwell, D. D.) (Geol. Soc. Am., Boulder, 1991).

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. The work of M.D.Z. and L.V. was supported by the National Science Foundation. German scientists were supported by the Ministry of Research and Technology and the German Science Foundation. We thank L. Liu, M. L. Zoback and S. Hickman for comments on the manuscript.

Global-scale latitudinal patterns of species diversity in the deep-sea benthos

Michael A. Rex*, Carol T. Stuart*, Robert R. Hessler†, John A. Allen‡, Howard L. Sanders§ & George D. F. Wilson||

* Department of Biology, University of Massachusetts, Boston, Massachusetts 02125, USA

† Scripps Institution of Oceanography, La Jolla, California 92093, USA

‡ University Marine Biological Station, Millport, Isle of Cumbrae, Scotland KA28 0EG, UK

§ Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

|| Australian Museum, Sydney South, New South Wales 2000, Australia

LATITUDINAL gradients of species diversity are ubiquitous features of terrestrial and coastal marine biotas, and they have inspired the development of theoretical ecology¹⁻³. Since the discovery of high species diversity in the deep-sea benthos⁴, much has been learned about local^{5,6} and regional⁷⁻⁹ patterns of diversity. Variation in diversity on larger scales remains poorly described. Latitudinal gradients of diversity were unexpected because it was assumed that the environmental gradients that cause large-scale patterns in surface environments could not affect communities living at great depths¹⁰. Here we report that deep-sea bivalves, gastropods and isopods show clear latitudinal diversity gradients in the North Atlantic, and strong interregional variation in the South Atlantic. Many seemingly incompatible mechanisms have been proposed to explain deep-sea species diversity¹¹. The existence of regular global patterns suggests that these mechanisms operate at different spatial and temporal scales.

The material analysed here represents the Gastropoda, Bivalvia and Isopoda from 97 epibenthic sled samples⁴ collected between 37° S in the Atlantic Ocean and 77° N in the Norwegian Sea (Fig. 1). All samples were taken in soft-sediment habitats at depths of 500–4,000 m, and in total comprised 214,508 individuals. This is the largest and most geographically extensive database to have been analysed for patterns of biodiversity in the deep sea. Data for the Atlantic originate from our research on the systematics and ecology of these three taxa. We used published relative abundance data for mollusks¹² and isopods¹³ in the Norwegian Sea. Species diversity was measured by using Hurlbert's¹⁴ modification of Sanders'⁷ rarefaction methodology to normalize diversity to a common number of individuals. For each taxon we used the largest normalized sample size possible to estimate diversity while still retaining geographic coverage sufficient to explore large-scale patterns.

All three taxa show highly significant latitudinal gradients of species diversity in the North Atlantic and Norwegian Sea (Fig. 2; Table 1). Elevated tropical diversity^{10,15} and depressed diversity in the Norwegian Sea^{13,16} have been noted earlier for deep-sea peracarid crustaceans, but the patterns presented here provide the first indication that continuous latitudinal gradients exist and are taxonomically widespread. The broad overlap in diversities between faunas of the Gambia and Guiana Basins (Fig. 2) shows that longitudinal bands of similar diversity can span the North Atlantic.

In the South Atlantic, comparable sampling of the deep-sea benthos has occurred over a much smaller latitudinal range, and far fewer samples are available (Figs 1 and 2). Bivalves and gastropods do show significant poleward declines in diversity, but the relationships are weaker than in the North Atlantic (Fig. 2; Table 1). Isopod diversity is not significantly related to latitude (Table 1) because of unusually high diversities in the Argentine Basin (Fig. 2). All three taxa show depressed diversity in

the Cape Basin off Walvis Bay, Namibia, and elevated diversity in the Argentine Basin. It is clear that the South Atlantic exhibits strong interregional variation and, in the case of isopods in the Argentine Basin, a centre of exceptionally high diversity at temperate latitudes. Additional sampling in the deep Southern Ocean beyond 40° S is needed to establish conclusively whether a general poleward reduction of diversity underlies the pronounced interregional variation.

In Fig. 2, much of the variation within regions reflects changes in diversity with depth across the bathyal zone^{8,9,13}. To test whether the latitudinal gradients shown in Fig. 2 could be spurious artefacts of differences in depth coverage among regions, we did a regression analysis of residuals to control for variation in depth (Fig. 3; Table 1). In the North Atlantic and Norwegian Sea, the relationships between diversity and latitude are still highly significant in all three taxa when depth is held constant. In the South Atlantic, relationships for mollusks are again significant, with the latitudinal trend for gastropods actually improving considerably, and the relationship for isopods remains insignificant.

Latitudinal diversity gradients in the deep-sea benthos of the North Atlantic and Norwegian Sea (Fig. 2) are very similar to latitudinal patterns in terrestrial and shallow marine biotas of the Northern Hemisphere^{1,3,17,18}. Diversity trends in surface biotas of the Southern Hemisphere are less well known, but

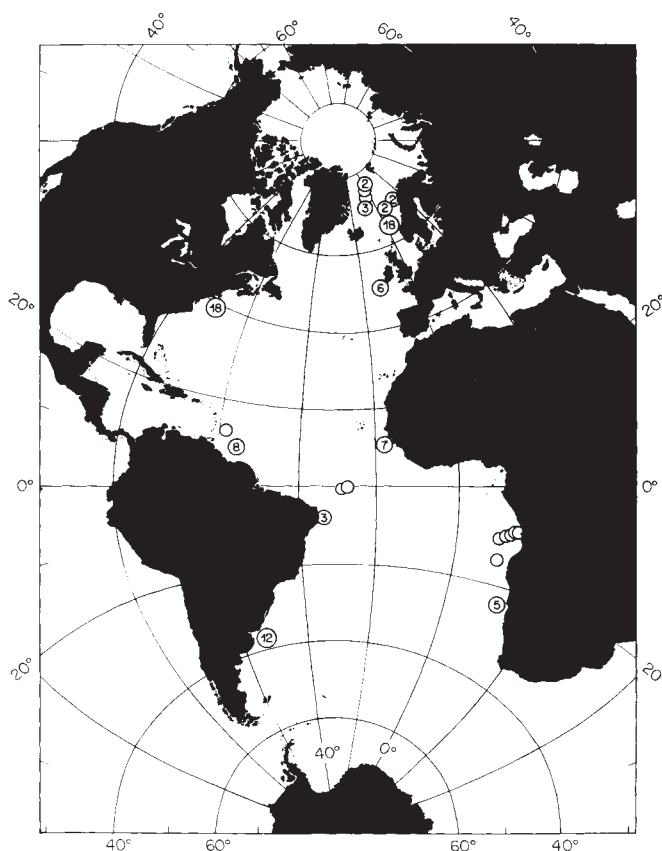


FIG. 1 Map of the Atlantic Ocean and Norwegian Sea showing the localities of the 97 epibenthic sled samples analysed in this study. Circled numbers indicate the number of samples taken in that region. Circles without numbers represent individual samples. The ten regions sampled from north to south are the Norwegian Sea, the West European Basin southwest of Ireland, the North American Basin south of New England, the Gambia Basin southwest of Senegal, the Guiana Basin off Surinam, the Equatorial Mid-Atlantic, the Brazil Basin off eastern Brazil, the Angola Basin off Angola, the Cape Basin off Namibia and the Argentine Basin off Uruguay and Argentina.

TABLE 1 Regression equations and their significance levels for the relationships of species diversity to latitude, and for these relationships with the effects of depth removed.

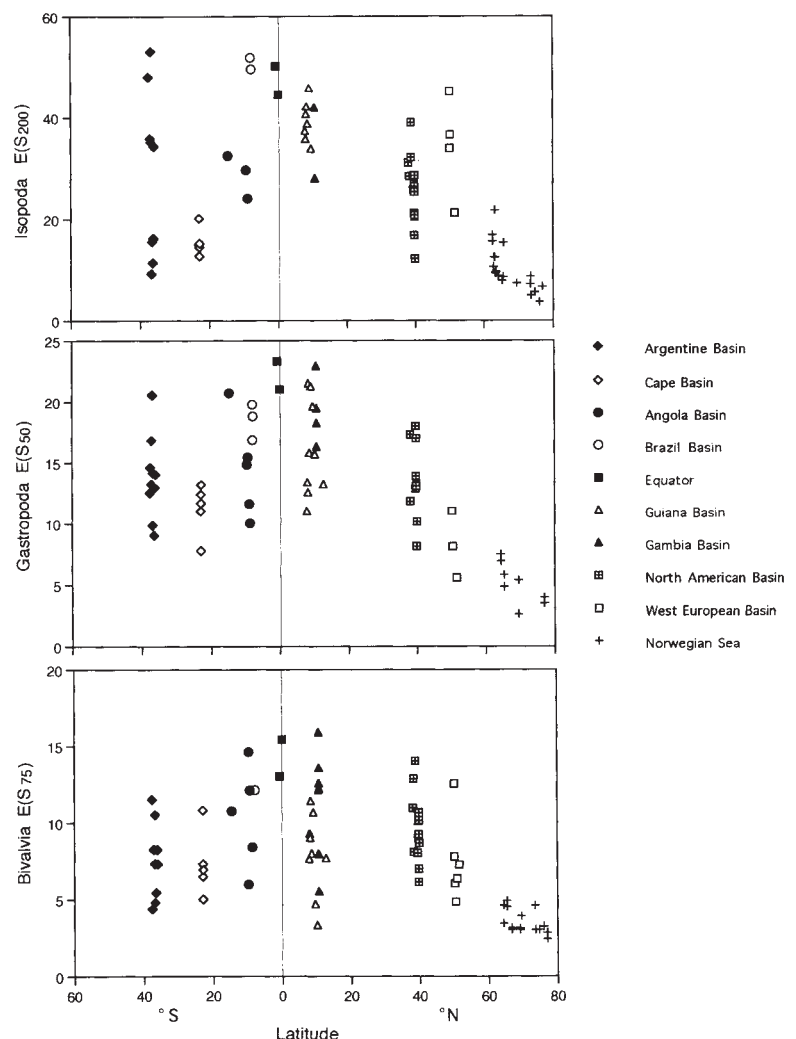
	North Atlantic		South Atlantic	
	Diversity versus latitude	Depth removed	Diversity versus latitude	Depth removed
Isopoda	$r = -0.865$ $F = 142.496$ $p = 0.0001, N = 50$ $Y = 45.029 - 0.490X$	$r = -0.871$ $F = 150.651$ $p = 0.0001, N = 50$ $Y = -0.489X$	$r = -0.352$ $F = 2.682$ $p = 0.1179, N = 21$	$r = -0.375$ $F = 3.103$ $p = 0.0942, N = 21$
Gastropoda	$r = -0.849$ $F = 90.708$ $p = 0.0001, N = 37$ $Y = 19.931 - 0.206X$	$r = -0.865$ $F = 104.400$ $p = 0.0001, N = 37$ $Y = -0.211X$	$r = -0.438$ $F = 5.683$ $p = 0.0254, N = 26$ $Y = 17.61 - 0.131X$	$r = -0.544$ $F = 10.084$ $p = 0.0041, N = 26$ $Y = -0.149X$
Bivalvia	$r = -0.687$ $F = 46.442$ $p = 0.0001, N = 54$ $Y = 11.782 - 0.102X$	$r = -0.672$ $F = 42.854$ $p = 0.0001, N = 54$ $Y = -0.101X$	$r = -0.600$ $F = 11.264$ $p = 0.0031, N = 22$ $Y = 12.296 - 0.145X$	$r = -0.606$ $F = 11.584$ $p = 0.0028, N = 22$ $Y = -0.144X$

Regression equations are not given for relationships that are not significant (Isopoda in the South Atlantic). North Atlantic includes the Norwegian Sea.

recent biogeographic and taxonomic research suggests a complex pattern of interregional variation with some groups having peak diversity at temperate rather than tropical latitudes^{19,20}. Latitudinal gradients in the Northern Hemisphere and more heterogeneous regionally specific variation in the Southern Hemisphere may be universal global patterns for both surface and deep-sea ecosystems.

It is important to recognize that species diversity has been measured somewhat differently in surface and deep-sea environments. Studies of large-scale diversity patterns in terrestrial and coastal marine habitats have focused primarily on estimates of the number of species that occur per unit area or across latitudinal bands^{1,3,17,18}. Data on the number of species are compiled typically from long-term, comprehensive and broadly based

FIG. 2 The relationships between species diversity and latitude for the Isopoda, Gastropoda and Bivalvia in the Atlantic Ocean and Norwegian Sea (see Fig. 1 for sampling localities). All samples are from bathyal depths (500–4,000 m). Diversity is calculated as Hurlbert's¹⁴ expected number of species $E(S_n)$, where n is the normalized sample size. Regression equations and significance levels for relationships of diversity to latitude are given in Table 1.



biotic surveys. By contrast, sampling methods to assess species diversity accurately in the deep-sea benthos have been developed only during the past 25 years¹, and the geographic coverage of sampling programs is much more limited. Patterns of diversity must be inferred from individual samples of deep-sea communities that are normalized to correct for variation in sample size. The normalized expected number of species $E(S_n)$ has been used extensively as a measure of deep-sea species diversity on local and regional scales^{6,8,9}. Using $E(S_n)$ here allows us to extend the analysis of deep-sea species diversity to global scales using a consistent measure of diversity. Like most diversity measures, the expected number of species $E(S_n)$ is affected by both the number of species and the relative abundance of species in samples²¹. However, $E(S_n)$ is an unbiased estimator of diversity which is sensitive to rare species^{22,23}. The actual number of species recovered in the samples is highly correlated with the normalized expected number of species for all three taxa shown in Fig. 2 (r values are 0.942, 0.570 and 0.844, $P=0.0001$ for isopods, gastropods and bivalves), and also reveals clear latitudinal gradients in the Northern Hemisphere (r values are -0.801 , -0.706 , -0.512 , $P=0.0001$, respectively).

We used Pielou's J (ref. 24) to determine the contribution of evenness of relative abundance to diversity. The pattern of evenness in northern samples varies among the three taxa. For bivalves, evenness is uncorrelated with latitude ($r=-0.168$, n.s.). Isopods and gastropods show significant decreases in evenness with latitude ($r=-0.571$, $P=0.0001$; $r=-0.379$, $P=0.0206$), but the relationships are weaker than for either $E(S_n)$ or the actual raw number of species (see Table 1, and above). Thus, latitudinal diversity gradients in both deep-sea and surface environments of the Northern Hemisphere reflect variation in the number of coexisting species, but diversity gradients in the deep sea also include a component of variation in evenness of relative abundance for some taxa.

Latitudinal diversity gradients are the most conspicuous biogeographical pattern in nature, and have been intensely studied. Nevertheless, their underlying causes remain uncertain. The numerous proposed explanations fall into two broad and inter-related categories: ecological and historical^{17,18,25}. Several lines of evidence suggest that both phenomena may be implicated at large scales in the deep sea.

The surprising congruity in global-scale patterns of diversity between surface and deep-sea biotas may result ultimately from ecological coupling through the water column. Recent experimental and observational research has demonstrated that sinking rates of organic particulates are much higher than expected, and are correlated with overhead surface production^{26,27}. The deep-sea benthos shows a physiological and numerical response to phytodetrital accumulation on the bottom^{28,29}. Latitudinal diversity gradients in the North Atlantic and Norwegian Sea (Fig. 2) correspond to poleward increases in surface production³⁰, particle flux to the deep sea³¹ and benthic biomass³². Carbon flux to the deep sea may be more variable as well at higher latitudes because surface production is more seasonal. The anomalously low diversities in the Cape Basin of the South Atlantic (Figs 1 and 2) are also associated with high and variable surface production³³. The relationship of nutrient input to community structure in the deep sea is apt to be indirect and to vary geographically³⁴. High and variable nutrient loading can depress diversity through several mechanisms, including the acceleration and destabilization of interspecific interactions⁸, and by altering the nature and scale of habitat heterogeneity⁹.

Historical factors also may be important. Local diversity in deep-sea prosobranch gastropods is a positive and significant function of regional diversity and the dispersal potential of the regional species pool³⁵. This suggests that local diversity is regulated by colonization from the regional species pool³⁶, and reflects the historical evolutionary build-up of regional diversity²⁵. The regional diversity of prosobranchs in deep-sea basins decreases from the equator to the Norwegian Sea. Low

diversity in the deep Norwegian Sea has been attributed to Quaternary glaciation¹⁶, during which extensive sea ice covered the Norwegian Sea and northern reaches of the Atlantic³⁷. Thus, the poleward decline in regional and consequently local diversity may represent, to some extent, a recovery from the effects of glaciation.

Patterns observed at global scales help to reconcile the myriad conflicting theories on deep-sea species diversity⁸. Much of the controversy about causes of diversity centres on the relative importance of probabilistic and deterministic mechanisms. Precision sampling studies⁶ and experimental manipulations³⁸ indicate that species coexistence on small scales (centimetres to metres) may be mediated through a mosaic of microhabitat patches created by random biogenic disturbance from the activities of megabenthic species, and from sinking organic material. Advocates of the disturbance-mosaic theory contend that it explains high diversity throughout the deep sea³⁹. However, at larger scales, definite regular patterns of species diversity emerge. Extensive sampling in the western North Atlantic has revealed bathymetric gradients in diversity that are correlated with sediment heterogeneity⁹, and are associated with predictable changes in the standing stock and trophic make-up of the benthos⁸. We have documented consistent latitudinal diversity gradients in the North Atlantic and Norwegian Sea, and interregional variation in the South Atlantic. Collectively, these findings suggest that the mechanisms that regulate diversity operate at different scales, just as they seem to in other marine and terrestrial systems⁴⁰.

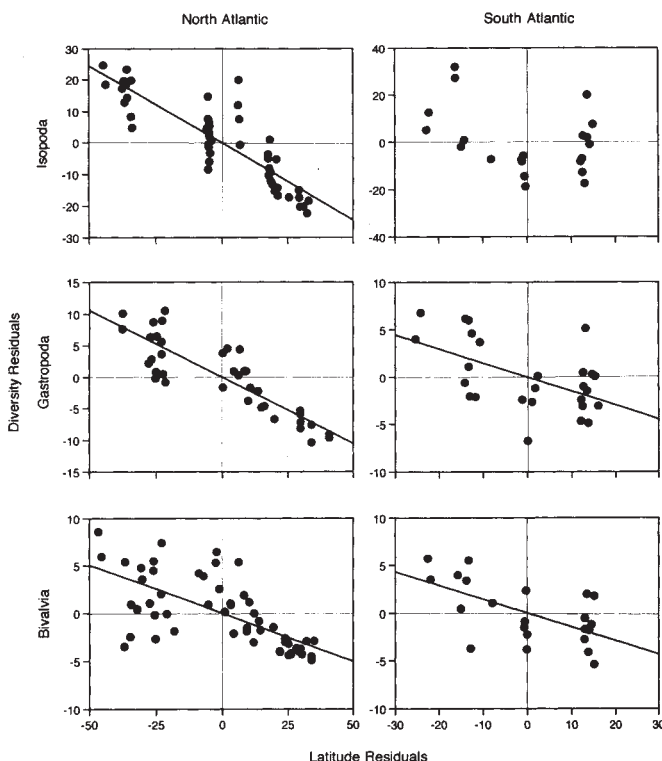


Fig. 3 The relationships of the residuals of diversity versus depth plotted against the residuals of latitude versus depth. Plots for the North Atlantic include data for the Norwegian Sea. These partial regression analyses remove the effect of depth from the relationship between diversity and latitude. As species diversity in the deep-sea benthos can vary with depth and geographically, it is important to hold depth constant to detect a latitudinal component to variation in diversity. The analyses show that variation in depth has no significant effect on latitudinal diversity patterns shown in Fig. 2 (see Table 1 for regression equations and their significance levels). No regression line is shown for Isopoda in the South Atlantic because the relationship is not significant (Table 1).

Stochastic patch dynamics are important in maintaining diversity at local scales. At larger regional and interregional scales, environmental gradients and the evolutionary history of community development become progressively more important in shaping patterns of diversity.

Large-scale latitudinal diversity gradients and interregional differences have important ramifications for assessing biodiversity for conservation purposes. The trends in normalized alpha diversity shown in Fig. 2 cannot in themselves provide estimates of global biodiversity for the deep-sea benthos. But these patterns demonstrate that attempts to project global biodiversity from the results of regionally based sampling programs⁶ must include the significant variation in diversity among deep-sea basins^{41,42}. Similarly, management decisions concerning exploitation of deep-sea environments for mining, petroleum exploration and waste disposal, and the potential impact of these activities on the deep-sea benthos, must recognize that patterns of community structure and their causes vary significantly with scale in this vast ecosystem. □

Received 1 March; accepted 12 August 1993.

1. Fischer, A. G. *Evolution* **14**, 64–81 (1960).
2. Pianka, E. R. *Am. Nat.* **100**, 33–46 (1966).
3. MacArthur, R. H. *Geographical Ecology: Patterns in the Distribution of Species* (Harper & Row, New York, 1972).
4. Hessler, R. R. & Sanders, H. L. *Deep-Sea Res.* **14**, 65–78 (1967).
5. Jumars, P. A. & Eckman, J. E. in *Deep-Sea Biology* (ed. Rowe, G. T.) 399–451 (Wiley, New York, 1983).
6. Grassle, J. F. & Maciolek, N. J. *Am. Nat.* **139**, 313–341 (1992).
7. Sanders, H. L. *Am. Nat.* **102**, 243–282 (1968).
8. Rex, M. A. in *Deep-Sea Biology* (ed. Rowe, G. T.) 453–472 (Wiley, New York, 1983).
9. Etter, R. J. & Grassle, J. F. *Nature* **360**, 576–578 (1992).
10. Jones, N. S. & Sanders, H. L. *Deep-Sea Res.* **19**, 737–745 (1972).
11. Gage, J. D. & Tyler, P. A. *Deep-Sea Biology: A Natural History of Organisms at the Deep-Sea Floor* (Cambridge Univ. Press, UK, 1991).

12. Bouchet, P. & Warén, A. *Sarsia* **64**, 211–243 (1979).
13. Svavarsson, J., Brattegard, T. & Strömberg, J. O. *Progr. Oceanog.* **24**, 297–310 (1990).
14. Hurlbert, S. H. *Ecology* **52**, 577–586 (1971).
15. Hessler, R. R. & Wilson, G. D. F. in *Evolution, Time and Space: The Emergence of the Biosphere* (eds Sims, R. W., Price, J. H. & Whalley, P. E. S.) 227–254 (Academic, New York, 1983).
16. Dahl, E. *Sarsia* **64**, 57–59 (1979).
17. Stevens, G. C. *Am. Nat.* **133**, 240–256 (1989).
18. Pagel, M. D., May, R. M. & Collie, A. R. *Am. Nat.* **137**, 791–815 (1991).
19. Clarke, A. *Trends Ecol. Evol.* **7**, 286–287 (1992).
20. Platnik, N. I. in *Systematics, Ecology and the Biodiversity Crisis* (ed. Eldredge, N.) 15–24 (Columbia Univ. Press, New York, 1992).
21. Magurran, A. E. *Ecological Diversity and its Measurement* (Princeton Univ. Press, New Jersey, 1988).
22. Smith, W. & Grassle, J. F. *Biometrics* **33**, 283–292 (1977).
23. Smith, W., Grassle, J. F. & Kravitz, D. in *Ecological Diversity in Theory and Practice* (eds Grassle, J. F., Patil, G. P., Smith, W. & Tailie, C.) 177–191 (International Coop., Maryland, 1979).
24. Pielou, E. C. *An Introduction to Mathematical Ecology* (Wiley, New York, 1969).
25. Ricklefs, R. E. *Science* **235**, 167–171 (1987).
26. Billet, D. S. M., Lampitt, R. S., Rice, A. L. & Mantoura, R. F. C. *Nature* **302**, 520–522 (1983).
27. Deuser, W. G. et al. *Deep-Sea Res.* **37**, 1331–1343 (1990).
28. Graf, G. *Nature* **341**, 437–439 (1989).
29. Lamshead, P. J. D. & Gooday, A. J. *Deep-Sea Res.* **37**, 1263–1283 (1990).
30. Campbell, J. W. & Aarup, T. *Deep-Sea Res.* **39**, 1669–1694 (1992).
31. Sibuet, M., Lambert, C. E., Chesselet, R. & Laubier, L. *J. Mar. Res.* **47**, 851–867 (1989).
32. Rowe, G. T. in *Deep-Sea Biology* (ed. Rowe, G. T.) 97–121 (Wiley, New York, 1983).
33. Sanders, H. L. *Brookhaven Symp. Biol.* **22**, 71–81 (1969).
34. Watts, M. C., Etter, R. J. & Rex, M. A. in *Deep-Sea Food Chains and the Global Carbon Cycle* (eds Rowe, G. T. & Pariente, V.) 245–254 (Kluwer, Netherlands, 1992).
35. Stuart, C. T. & Rex, M. A. in *Reproduction, Larval Biology and Recruitment in the Deep-Sea Benthos* (eds Young, C. M. & Eckelbarger, K. J.) (Columbia Univ. Press, New York, in the press).
36. Cornell, H. V. *Am. Nat.* **126**, 565–569 (1985).
37. Dunton, K. *Trends Ecol. Evol.* **7**, 183–189 (1992).
38. Kukert, H. & Smith, C. R. *Deep-Sea Res.* **39**, 1349–1371 (1992).
39. Grassle, J. F. *Trends Ecol. Evol.* **4**, 12–15 (1989).
40. Levin, S. A. *Ecology* **73**, 1943–1967 (1992).
41. May, R. M. *Nature* **357**, 278–279 (1992).
42. May, R. M. *Nature* **361**, 598 (1993).

ACKNOWLEDGEMENTS. We thank R. Etter, W. Ponder and A. Rex for reading the manuscript, and the National Science Foundation for supporting this research. W. Dunkle of the Data Library and Archives at the Woods Hole Oceanographic Institution provided the map used in Fig. 1.

A selfish strategy of social insect workers that promotes social cohesion

David C. Queller, Joan E. Strassmann,
Carlos R. Solís, Colln R. Hughes*
& Debbie Moralez DeLoach

Department of Ecology and Evolution, Rice University, PO Box 1892, Houston, Texas 77251, USA

SOCIAL insect colonies are in many ways analogous to organisms¹, because kin-selected workers act selflessly and cohesively to promote the fitness of a few reproductive members^{2,3}. But workers can evolve selfish strategies which create reproductive conflict, reducing the functional integrity of colonies. For example, they can lay unfertilized (male) eggs^{3–6}, compete directly with the queen to lay fertilized (female) eggs⁷, suppress the reproduction of other workers^{8,9}, choose among several queens¹⁰ and generally favour closer over more distant kin¹¹. Conflicts over the sex ratio may be especially pervasive, even in highly eusocial insects. The unusually high relatedness ($r = 3/4$) of female hymenopteran workers to their full sisters means that workers should prefer more female-biased sex ratios than do queens⁴. The worker preference for females should be exerted most strongly on colonies where they are most likely to be full sisters, leaving male production to colonies where this advantage least applies^{4,12–14}; this prediction is supported by studies in ants and bees^{12,15–17}. Here we show that when colonies have multiple queens born in the same nest, the selfish worker

sex-ratio strategy has a paradoxical side-effect which strongly promotes social cohesion. This strategy accounts for the peculiar colony cycle of epiponine wasps, and may be responsible for the maintenance of eusociality in this group.

When colonies differ in queen number, those with the fewest queens will have the most full-sister relationships, so workers in these colonies are predicted to rear females. Males should be reared in colonies with larger numbers of queens (Fig. 1). When queens are natal nestmates, this selfish worker strategy has an unnoticed side-effect that will stabilize sociality, by keeping relatedness from falling to very low levels. In haplodiploids, relatedness among the daughters of N equally fecund queens related by r_q is

$$r = \frac{3}{4N} + \frac{r_q}{4} \left(1 - \frac{1}{N} \right) \quad (1)$$

assuming each queen mates singly¹⁸ (when N varies, the harmonic mean should be used¹⁹). Now, let r be the relatedness among a special set of progeny, cohorts of newly produced queens. At equilibrium $r_q = r$, and both approach zero for large N . But in this application, N is the number of old queens in the colonies producing new queens. If workers rear new queens only in the few colonies with one old queen, then $r_q = 3/4$ and relatedness among their female progeny always remains above $3/16$, even in colonies with many queens. Thus, by following a selfish sex-ratio strategy, workers incidentally keep future relatednesses from falling to levels that would make kin selection ineffective and destabilize sociality.

Epiponine wasps²⁰ are a tribe of neotropical vespids that meet the assumptions of the model; they have multiple queens (Fig. 2a) who are related because new colonies arise through fissioning^{21–23}. Hamilton³ viewed these wasps as a serious challenge to kin-selection theory. How could sociality be stable if high queen numbers make relatedness very low, particularly

* Present address: Department of Biology, University of North Dakota, Box 8238, University Station, Grand Forks, North Dakota 58202, USA.

TABLE 1 Relatedness among adult colony-mates for the four epiponine populations of Fig. 1

Type of relatedness	Relatedness $\pm$ s.e. (no. of colonies, no. of individuals)			
	<i>Polybia occidentalis</i>	<i>Protopolybia exigua</i>	<i>Polybia emaciata</i>	<i>Parachartergus colobopterus</i>
Queens	0.58 $\pm$ 0.08 (27,216)	0.82 $\pm$ 0.07 (29,223)	0.55 $\pm$ 0.10 (20,193)	0.66 $\pm$ 0.09 (19,156)
Workers in colonies with males	0.27 $\pm$ 0.07 (13,159)	0.49 $\pm$ 0.13 (13,107)	0.31 $\pm$ 0.10 (12,132)	0.30 $\pm$ 0.15 (8,116)

Relatednesses were estimated from allozyme data (see ref. 25 for methods) according to ref. 32 using the computer program Relatedness 4.1. Colonies were weighted equally. Standard errors (s.e.) were estimated by jack-knifing over colonies, converted to variances and used in one-tailed *t*-tests (unequal variances) of relatedness differences. As predicted by the worker control hypothesis, relatedness among queens is always significantly higher than relatedness among workers in male-producing nests ($P < 0.05$).

when workers are not locked into their role by a specialized morphology²¹⁻²³? However, estimated progeny relatednesses are unexpectedly high, typically ~ 0.3 , owing to the fact that queens are very closely related, often as full sisters^{18,24-27} (Table 1). The high queen relatedness stems from a peculiar colony cycle²⁸⁻³¹ (Fig. 2b). Each colony begins with many queens, and at any given time most colonies have multiple queens (Fig. 2a), but new queens are reared only in those colonies where queen number has been reduced to almost one by attrition or social competition.

This pattern is predicted under worker control of sex ratios, but it could also be caused by queen politics or by the need for colony growth before reproduction. A more definitive test is provided by the timing of male production, which has received much less study. If males are also produced on colonies with very few queens, there is simply a general reproductive delay. Worker control is implicated only if males are produced on colonies with comparatively large numbers of queens.

Queen production is both rare and cryptic (owing to the external similarity of workers and queens), making such events hard to study. But the relatedness among the current generation of

queen colony-mates can be used as an index of the number of old queens that produced them. A comparable index for the number of queens on male-producing colonies is the relatedness among female workers on those colonies. If this relatedness is as high as queen relatedness, then the worker-control prediction that there are more queens on male-producing colonies fails. But if it is lower, the worker-control prediction is supported.

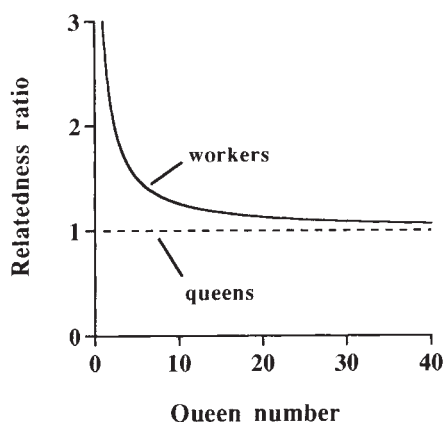


FIG. 1 The effect of queen number on the ratio (relatedness to new queens)/(relatedness to males). The ratios plotted here include a reproductive-value devaluation of males by 1/2 (due to haploidy). For workers the ratio decreases with increasing number of old queens, because there are fewer full sisters with the crucial 3/4 relatedness. Therefore, workers should be selected to favour production of females in colonies with few queens and production of males in colonies with comparatively many queens^{13,14}. Queens have no such preference. The two strategies can therefore be distinguished without attempting to predict or measure the population sex investment ratio (which would be complicated by colony fissioning). The functions shown assume equal egg-laying by singly mated queens, who are related by 3/4. Given that queens are related, the general trends are unchanged unless either mate number or queen relatedness varies strongly and systematically with queen number. Such systematic variation is unlikely, particularly for species in which near-maximal queen relatedness leaves little room for any variation in either queen relatedness (uniformly high) or mate number (usually 1).

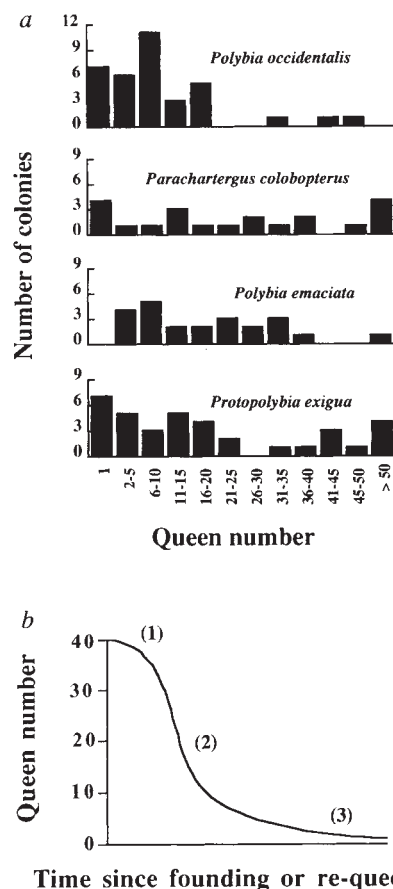
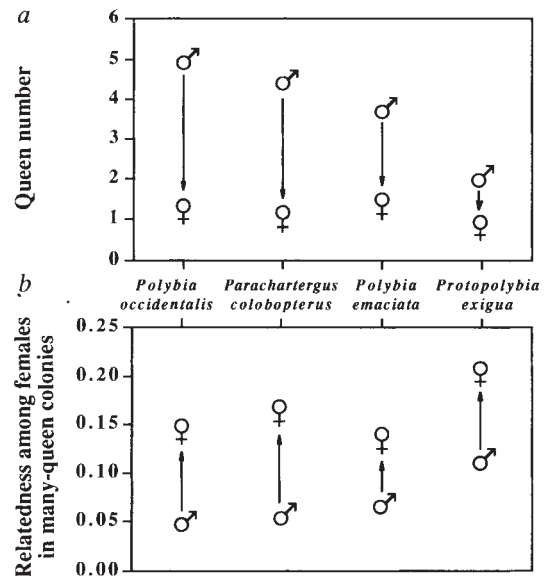


FIG. 2 Queen numbers in epiponine wasps. a, Distributions of queen numbers in four species. Entire colonies were collected from local populations in Venezuela. Queenship was determined by dissection (presence of mature eggs). Harmonic means for the four species, from top to bottom, are 3.1, 4.9, 6.7 and 3.7. For methods and additional data see refs 25-27. b, The queen-number cycle²⁸⁻³¹. Colonies begin with many queens (1), and then go through an extended period of decline in queen number (2), often ending with a single queen (3). Production of new queens, often followed by colony fission, occurs only in colonies at or approaching stage 3. The timing of male production is poorly known, but is predicted to be at stages 1 or 2 if workers control sex ratios (Fig. 1). Males emigrate and mate with queens from other colonies.

FIG. 3 Effects of worker sex-ratio preferences. *a*, Effective queen numbers inferred for male-producing (male symbol) and queen-producing (female symbol) colonies. Arrows indicate the decline in queen number between rearing of males and new queens. Effective queen number is the harmonic mean number of equally reproducing queens required to produce the observed progeny relatedness. These are calculated by solving equation (1) for N , using progeny r values from Table 1 (row 2 for male-producing colonies, row 1 for queen-producing colonies). The assumption that queens mate once is supported by the very high queen relatednesses of at least two species and by the fact that using the assumption generates accurate estimates of worker relatedness^{25, 27}. Multiple mating would lower all queen number estimates. *b*, Relatedness among female progeny on colonies with many queens under two-queen production scenarios. Relatedness is calculated as $r_q/4$ (equation (1) for asymptotically high queen number N). The female values are calculated using the empirical estimates of r_q from Table 1. The values labelled by the male symbol use the value of r_q that would apply if new queen cohorts were produced instead on the colonies that actually produced males. To obtain this hypothetical r_q , we use equation (1), letting N be the number of old queens in male-producing colonies shown in *a*, and solving for r_q under the equilibrium condition that $r_q = r$ (relatedness among parent queens equals relatedness among progeny queens). Arrows indicate how much relatedness is raised as a side-effect of the delay in queen production. The female progeny on many-queen colonies are normally workers, but similar patterns are obtained for relatednesses of these females to the new queens and males that will eventually be reared.



Relatednesses for four species of epiponine wasps are shown in Table 1; all four support the worker-control prediction. Figure 3*a* shows the inferred queen-number differences between male-producing and queen-producing colonies.

The use of relatedness to infer relative queen numbers provides an indirect but robust test of the prediction. That new queens are produced when old queens are few is corroborated by behavioural observations of several species (including *Polybia occidentalis*)^{28, 31}. The lower relatedness of workers in male-producing colonies clearly establishes that there are more queens on these colonies. Even if only one of these queens produced the reproductive members, the mixed maternity of workers ensures that average relatedness of workers to female reproductive members would be well below 3/4. These colonies would still be predicted to rear males because the workers could not rear many full sisters. The prediction also remains unchanged if workers lay unfertilized eggs that become males (which has not been observed). The workers who stand to gain most from rearing females are still those in colonies with the fewest queens.

Alternative explanations for the pattern seem unlikely. Colonies might specialize in different sexes based on differences in the degree of local mate competition or local resource competition, but these factors are expected to increase with colony size, and there are no significant differences in worker number between

colonies that produce or do not produce males (*t*-tests, $P > 0.15$). The re-queening part of the pattern could be explained trivially: queen number rises after re-queening, so it necessarily occurs at times with lower queen number. Re-queening at low queen numbers could also be an adaptive response to the need for new queens to keep the colony going. However, neither of these explanations accounts for the similar timing for production of new queens that will found new colonies rather than staying in the old one.

If selfish worker sex-ratio strategies account for the peculiar epiponine colony cycle, then they are also indirectly responsible for the surprisingly high progeny relatednesses observed, and perhaps for the maintenance of sociality. The effect is strongest for colonies at the crucial many-queen stage (high N), when relatedness would otherwise fall to very low values. For example, if queens were produced in the colonies that actually produce males, then relatedness among daughters of many-queen colonies in these species would be as low as 0.05 (Fig. 3*b*, male symbols). The actual colony cycle results in a two- to threefold increase in relatedness at this stage (Fig. 3*b*, female symbols).

Thus, although selfish worker strategies may typically disrupt colony cohesion, worker sex-ratio strategies can produce an opposite effect in multi-queen species, and this may be a crucial factor in maintaining sociality in epiponine wasps. □

Received 2 April; accepted 12 July 1993.

1. Wheeler, W. M. *J. Morph.* **22**, 307–325 (1911).
2. Hamilton, W. D. *J. theor. Biol.* **7**, 1–16, 17–52 (1964).
3. Hamilton, W. D. *Rev. ecol. Syst.* **3**, 193–232 (1972).
4. Trivers, R. L. & Hare, H. *Science* **191**, 249–263 (1976).
5. Bourke, A. F. G. *Q. Rev. Biol.* **63**, 291–311 (1988).
6. Choe, J. C. in *Advances in Myrmecology* (ed. Trager, J. C.) 163–187 (Brill, Leiden, 1988).
7. Sommer, K. & Hölldobler, B. *Naturwissenschaften* **79**, 470–472 (1992).
8. Starr, C. K. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 427–464 (Academic, Orlando, 1984).
9. Ratnieks, F. L. W. & Visscher, P. K. *Nature* **342**, 796–797 (1989).
10. Forsyth, A. *Am. Nat.* **116**, 895–898 (1980).
11. Getz, W. in *Kin Recognition* (ed. Hepper, P.) 358–412 (Cambridge Univ. Press, UK, 1991).
12. Boomsma, J. J. & Grafen, A. *Evolution* **44**, 1026–1034 (1990).
13. Boomsma, J. J. & Grafen, A. *J. evol. Biol.* **3**, 383–407 (1991).
14. Pamilo, P. *Am. Nat.* **137**, 83–107 (1991).
15. Ward, P. S. *Behav. Ecol. Sociobiol.* **12**, 301–307 (1983).
16. Boomsma, J. J. *Trends ecol. Evol.* **6**, 92–95 (1991).
17. Mueller, U. G. *Science* **254**, 442–444 (1991).
18. Queller, D. C., Strassmann, J. E. & Hughes, C. R. *Science* **242**, 1155–1157 (1988).
19. Wade, M. J. *J. theor. Biol.* **112**, 109–121 (1985).
20. Carpenter, J. in *Biological Relationships Between Africa and South America* (ed. Goldblatt, P.) (Yale Univ. Press, New Haven, in the press).

21. Richards, O. W. *The Social Wasps of the Americas Excluding the Vespinae* (British Museum (Natural History), London, 1978).
22. Jeanne, R. L. *A. Rev. Ent.* **25**, 371–396 (1980).
23. Jeanne, R. L. in *The Social Biology of Wasps* (eds Ross, K. & Matthews, R.) 191–231 (Cornell Univ. Press, Ithaca, New York, 1991).
24. West-Eberhard, M. J. in *Social Insects in the Environment: Proc. 11th int. Congr. IUSSI* (eds Veeresh, G. K., Mallik, B. & Viraktamath, C. A.) 254–255 (Oxford and IBH, New Delhi, 1990).
25. Strassmann, J. E. et al. *Anim. Behav.* **42**, 461–470 (1991).
26. Queller, D. C. et al. *Behav. Ecol.* **4**, 7–13 (1993).
27. Gastreich, K. R., Strassmann, J. E. & Queller, D. C. *Ecol. Ethol. Evol.* (in the press).
28. West-Eberhard, M. J. in *Proc. 7th int. Congr. IUSSI* 396–403 (1973).
29. West-Eberhard, M. J. *Science* **200**, 441–443 (1978).
30. West-Eberhard, M. J. in *Natural Selection and Social Behaviour* (eds Alexander, R. D. & Tinkle, D. W.) 3–17 (Chiron, New York, 1981).
31. Forsyth, A. B. thesis, Harvard Univ. (1978).
32. Queller, D. C. & Goodnight, K. F. *Evolution* **43**, 258–275 (1989).

ACKNOWLEDGEMENTS. We thank J. Gonzalez, M. A. Castillo H., A. Fernandez Badillo, T. Blohm, the Arp, Swain and Castillo families, Universidad Central de Venezuela (Maracay) and FUSAGRI (Cagua), for access to collecting sites. Rice undergraduates J. A. Negrón-Sotomayor, K. R. Gastreich, E. Aviles, R. Ocanas and A. Treves assisted in data collection. K. F. Goodnight supplied the computer program Relatedness 4.1. This work was supported by the US NSF.

Induction of relapsing paralysis in experimental autoimmune encephalomyelitis by bacterial superantigen

Stefan Brocke^{*†}, Amitabh Gaur[‡],
Christopher Piercy^{*}, Anand Gautam^{*},
Koenraad Gijbels^{*}, C. Garrison Fathman[‡]
& Lawrence Steinman^{*}

Departments of ^{*}Neurology and Neurological Sciences and

[‡]Medicine, Stanford University School of Medicine,
Stanford, California 94305, USA

THE role of infection in the pathogenesis of clinical relapses that occur in most autoimmune diseases, including multiple sclerosis, remains to be established^{1,2}. Experimental autoimmune encephalomyelitis (EAE) serves as a model for multiple sclerosis, with episodes of relapsing paralysis³⁻⁹. In certain strains of mice, T-lymphocytes expressing the V β 8 T-cell receptor (TCR)⁶⁻⁸ engage the amino-terminal epitope Ac1-11 of myelin basic protein, leading to EAE. The bacterial superantigen staphylococcal enterotoxin B (SEB) activates V β 8-expressing T cells. Here we show that after immunization with Ac1-11, or after transfer of encephalitogenic T-cell lines or clones reactive to Ac1-11, SEB induces exacerbation or relapses of paralytic disease in mice that are in clinical remission following an initial episode of paralysis, and triggers paralysis in mice with subclinical disease. Tumour necrosis factor has a critical role in the mechanism underlying SEB-induced exacerbation of disease, because anti-tumour necrosis factor antibody given *in vivo* delays the onset of paralysis triggered by SEB. On reactivation of autoaggressive cells through their T-cell receptor, superantigens may induce clinical relapses of autoimmune disease.

We examined the effect of SEB on (PL/J $\times$ SJL) F1 mice previously immunized with the encephalitogenic MBP peptide Ac1-11. The effect of SEB was tested as well in PL/J and SJL/J mice after immunization with spinal cord homogenate. Mice were given 50 μ g of either SEB or staphylococcal enterotoxin A (SEA) intraperitoneally (i.p.) 21-30 days after immunization with Ac1-11, at a time when most mice had recovered from the initial paralytic disease. Data in Table 1 show the increase in the frequency of mice with disease exacerbation after treatment with SEB, SEA or phosphate-buffered saline (PBS).

In the PBS control group, few relapses were observed (14-20%; Table 1). In comparison, a significant increase (P -values <0.05 - 0.001) in numbers of mice with relapses of disease was recorded in the SEB-treated group within 14 days (67-86%; Table 1). The increase in relapses in the group treated with SEA was between 38-43% (Table 1), not statistically different to the rate in the PBS-treated group ($P>0.05$, SEA versus PBS). SEA could be acting on encephalitogenic T cells that are not V β 8⁺ (refs 7, 8, 10-12). A similar SEB-induced increase in disease exacerbation was observed in PL/J mice (Table 1; $P<0.01$, SEB versus PBS).

It was also possible to induce EAE in mice that were not paralysed after initial immunization with Ac1-11. In mice that failed to develop signs of EAE, 6 out of 10 developed signs of EAE when injected with SEB, compared with only 2 of 11 PBS-treated mice. These data suggest that SEB can exacerbate subclinical-EAE. But SEB did not worsen EAE in SJL/J mice, which develop EAE though V β 8.2⁺ T cells are deleted in this strain (Table 1)¹³. This implies that SEB activates pathogenic V β 8⁺ T cells to exacerbate EAE. A nonspecific effect of SEB on other cells *in vivo* is unlikely. SJL/J mice possess T-cell populations

TABLE 1 Percentage of mice with increase in EAE scores after superantigen treatment

	Mouse strain /H-2	SEB	SEA	PBS
Expt 1:	(PL/J $\times$ SJL) F1 /H-2 ^{u x s}	6/9 (67%)*	3/8 (38%)†	2/10 (20%)*†
Expt 2:	(PL/J $\times$ SJL) F1 /H-2 ^{u x s}	12/14 (86%)*	ND	2/13 (15%)*
Expt 3:	(PL/J $\times$ SJL) F1 /H-2 ^{u x s}	13/16 (81%)*	6/14 (43%)†	1/7 (14%)*†
Expt 4:	PL/J /H-2 ^u	5/10 (50%)*	ND	0/10 (0%)*
Expt 5:	SJL/J /H-2 ^s	1/18 (6%)	ND	0/6 (0%)

Effect of superantigen treatment of (PL/J $\times$ SJL) F1, PL/J or SJL/J mice on relapsing EAE. Groups of mice were immunized with MBP peptide Ac1-11 (experiments 1-3) or spinal cord homogenate (experiments 4 and 5) and PTX. Three to four weeks later, when most mice had recovered from EAE, mice were challenged with 50 μ g SEB, 50 μ g SEA or PBS i.p. Disease scores were recorded daily. The number of mice showing an increase in the mean disease scores up to 14 days after SEB, SEA or PBS administration out of the total number of mice in each group are shown in the table. Two- to three-month-old (PL/J $\times$ SJL) F1, PL/J or SJL/J mice purchased from Jackson Laboratory (Bar Harbor, ME) were immunized by subcutaneous inoculation at the base of the tail with 200 μ g of the MBP peptide Ac1-11 or 50 mg lyophilized spinal cord homogenate in an 0.1 ml emulsion with equal volumes of PBS and complete Freund's adjuvant (CFA; Difco, Detroit). Killed *Mycobacterium tuberculosis*, strain H37RA (Difco) was added to the emulsion at a concentration of 4 mg ml⁻¹. *Bordetella pertussis* toxin (200 ng; PTX, List Pharmaceuticals, Campbell, CA) was given i.v. on the day of peptide inoculation and 48 h later. All animals were followed for clinical signs of disease according to the following scale: 0, no clinical disease; 1, tail weakness; 2, paraparesis; 3, paraplegia; 4, paraplegia with forelimb weakness or paralysis; 5, moribund or dead animals. MBP peptide Ac 1-11 was prepared by continuous flow solid-phase synthesis according to the sequences for rat MBP (Ac-ASQKRPSQRHG) by the Protein and Nucleic Acid Facility, Beckman Center, Stanford University. Peptide purity was examined by high-performance liquid chromatography and peptide identity was confirmed by amino-acid composition analysis. SEB and SEA were purchased from Sigma (St Louis, MO) or Toxin Technology (Sarasota, FL) and injected in a concentration of 50 μ g in 500 μ l PBS i.p. per mouse. The difference in EAE score increases between the groups of mice treated with SEB and PBS were significant in H-2^u or H-2^{u x s} mice (*expt 1: $\chi^2=4.2$, $P<0.05$; expt 2: $\chi^2=13.4$, $P<0.001$; expt 3: $\chi^2=9.9$, $P<0.01$; expt 4: $\chi^2=6.7$, $P<0.01$), in contrast to disease score increases between PBS and SEA treated mice (†expts 1 and 3, $P>0.05$). ND, not determined.

like V β 7 that are activated by SEB, but nevertheless did not develop recurrence of EAE after a challenge with SEB.

To determine whether SEB alone could induce EAE, we injected 50-100 μ g SEB or SEA i.p. into naive PL/J or (PL/J $\times$ SJL) F1 mice (over 200 mice tested). No clinical sign of EAE followed administration of SEB or SEA in the absence of a previous antigen challenge with Ac1-11.

Pretreatment with SEB reduced the primary proliferative responses to myelin basic protein peptide Ac1-11, and decreased the severity of EAE^{12,14}. Similarly, superantigen protects rats against EAE¹⁵. The time of administration of SEB in the studies reported here produced different pathophysiological consequences from those in which SEB was given before immunization¹². In the first several days after SEB administration, a marked proliferation of T cells bearing the V β 8 T-cell receptor (TCR) occurs¹⁶⁻¹⁹. This initial proliferation correlates in time with disease reactivation induced with SEB, seen in mice in clinical remission after disease induction with Ac1-11 (Table 1). This proliferation could also account for disease exacerbation induced with SEB in mice immunized with Ac1-11 that had not exhibited clinical EAE (Table 1). But 1-2 weeks after SEB administration, the V β 8⁺ cells were rendered anergic^{12,16-19}. This

† To whom correspondence should be addressed.

could account for the protective effect of SEB two weeks before challenge with Ac1-11 (refs 12, 14).

To support further the hypothesis that SEB activates $V\beta 8^+$ encephalitogenic T cells, and is not simply 'inflammatory', we designed additional experiments using encephalitogenic and non-encephalitogenic T-cell lines and clones specific for Ac1-11. All clones and the majority of the line cells used the TCR $V\beta 8$ element (data not shown). Administration of SEB after transfer of 5×10^6 cells of the encephalitogenic clone TCMSB-C48 or the encephalitogenic line TLMSB-L9 leads to lethal EAE in all mice (Fig. 1a, b). Death occurs earlier with the clone C-48 (day 5), than with the line L-9 (days 7–11), partly because all of the cloned cells were $V\beta 8^+$.

The ability of Ac1-11 T-cell clones to induce EAE correlates with their production of TNF^{20} . *In vivo* administration of anti-TNF monoclonal antibody delayed the onset of the disease triggered by SEB. Two groups of 10 mice were given 50 μ g SEB

intravenously 4 days after receiving 5×10^6 cells of the Ac1-11 specific line TLMSB-L9. Mice given 1 mg anti-TNF antibody, XT-22, on day 3 after transfer, developed disease at 16.3 ± 1.2 days (mean $\pm$ s.d.), compared to 12.5 ± 0.7 days when given 1 mg of an isotype-matched control antibody, GL113 ($P < 0.01$, Mann-Whitney test).

EAE is not seen in animals when SEB is given after transfer of 5 – 10×10^6 cells of two non-encephalitogenic clones (TCMSB-C1, -C33B; Fig. 1c) or a non-encephalitogenic line (TLMSB-L1; Fig. 1c). Thus, the effect of SEB *in vivo* is dependent on the encephalitogenicity of TCR $V\beta 8^+$ T cells. The data suggest that the action of superantigen in exacerbating EAE is mediated by activation of encephalitogenic T cells, rather than nonspecific systemic cytokine release.

To test whether reinduction of EAE depends directly on stimulation of pathogenic T cells, or on systemic activation of T cells *in vivo* with subsequent cytokine release, we activated T

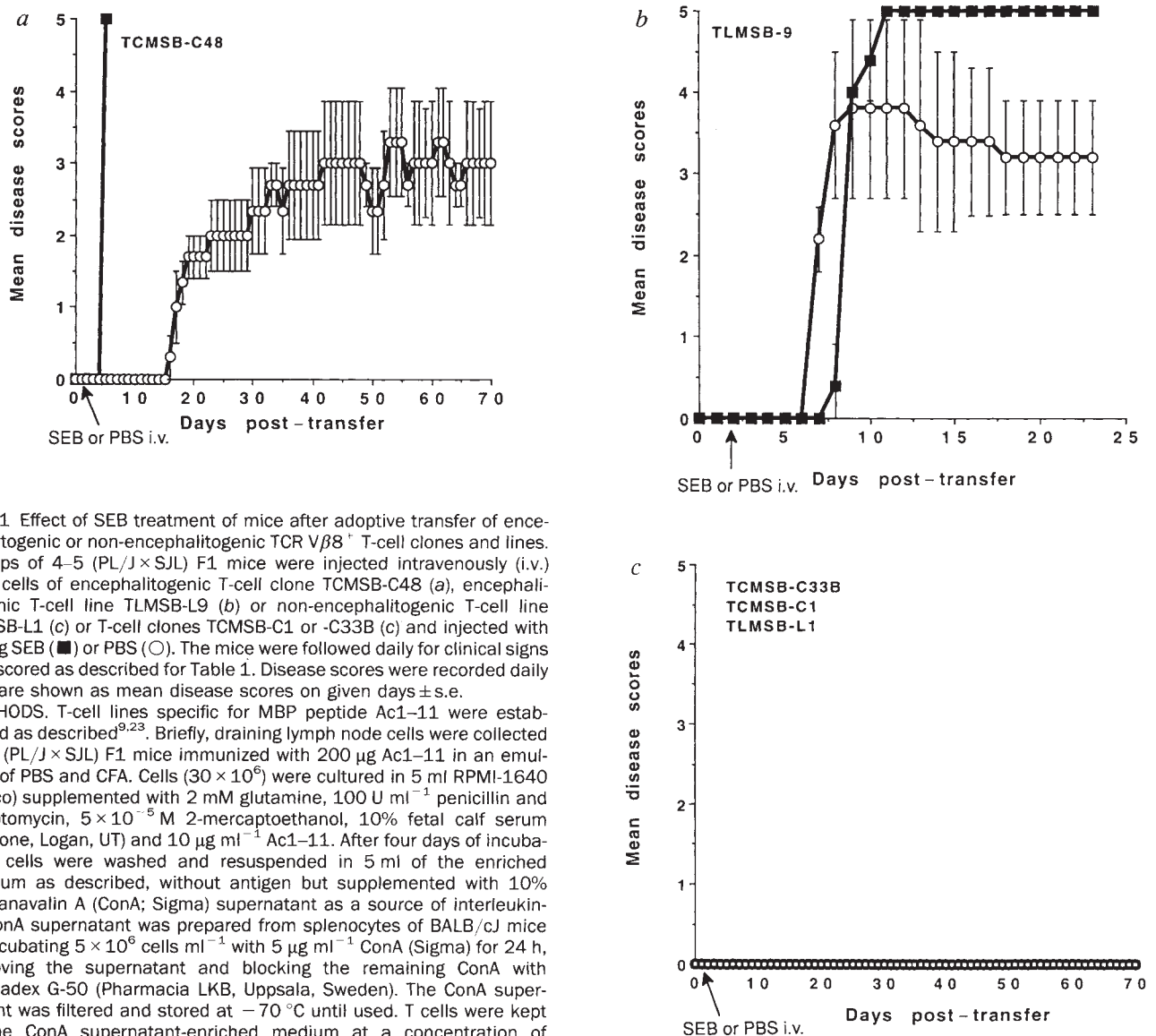


FIG. 1 Effect of SEB treatment of mice after adoptive transfer of encephalitogenic or non-encephalitogenic TCR $V\beta 8^+$ T-cell clones and lines. Groups of 4–5 (PL/J $\times$ SJL) F1 mice were injected intravenously (i.v.) with cells of encephalitogenic T-cell clone TCMSB-C48 (a), encephalitogenic T-cell line TLMSB-L9 (b) or non-encephalitogenic T-cell line TLMSB-L1 (c) or T-cell clones TCMSB-C1 or -C33B (c) and injected with 50 μ g SEB (■) or PBS (○). The mice were followed daily for clinical signs and scored as described for Table 1. Disease scores were recorded daily and are shown as mean disease scores on given days $\pm$ s.e.

METHODS. T-cell lines specific for MBP peptide Ac1-11 were established as described^{9,23}. Briefly, draining lymph node cells were collected from (PL/J $\times$ SJL) F1 mice immunized with 200 μ g Ac1-11 in an emulsion of PBS and CFA. Cells (30×10^6) were cultured in 5 ml RPMI-1640 (Gibco) supplemented with 2 mM glutamine, 100 U ml^{-1} penicillin and streptomycin, 5×10^{-5} M 2-mercaptoethanol, 10% fetal calf serum (Hyclone, Logan, UT) and 10 μ g ml^{-1} Ac1-11. After four days of incubation, cells were washed and resuspended in 5 ml of the enriched medium as described, without antigen but supplemented with 10% concanavalin A (ConA; Sigma) supernatant as a source of interleukin-2. ConA supernatant was prepared from splenocytes of BALB/c mice by incubating 5×10^6 cells ml^{-1} with 5 μ g ml^{-1} ConA (Sigma) for 24 h, removing the supernatant and blocking the remaining ConA with Sephadex G-50 (Pharmacia LKB, Uppsala, Sweden). The ConA supernatant was filtered and stored at $-70^\circ C$ until used. T cells were kept in the ConA supernatant-enriched medium at a concentration of 1×10^6 ml^{-1} and restimulated every 14 days using 10 μ g Ac1-11 ml^{-1} as antigen presented on irradiated syngeneic spleen cells (3,000 rads). T-cell clones from the Ac1-11-induced T-cell lines were derived by the limiting dilution technique. Cells were diluted in medium and distributed in wells of flat-bottom microtitre plates at a concentration of 0.2 cells per 200 μ l. Resulting clones were maintained by techniques described

for the lines. For adoptive transfer, 3 – 5×10^6 cells of the lines or clones were injected i.v. in 500 μ l PBS into naive mice after extensive washing. Before the transfer of clones, mice were irradiated (350 rads). Mice were challenged intraperitoneally with 50 μ g SEB in 500 μ l PBS or PBS as indicated.

TABLE 2 Transfer of EAE by TCR $V\beta 8^+$ T-cell line TLMSB-L9 or clone TCMSB-C1 after activation *in vitro* with SEB or Ac1-11

<i>In vitro</i> stimulation with	Ratio of IL-2R mean fluorescence intensity	Transferred cells	Number of diseased mice/total mice
(a) Experimental group			
Ac1-11	22.9	5×10^6 L9	10/10
SEB	20.5	5×10^6 L9	9/10
No activation	1.0	5×10^6 L9	0/10
(b) Controls			
SEB	ND	5×10^6 C1	0/5
Ac1-11	ND	5×10^6 C1	0/10
SEB	ND	5×10^6 control splenocytes	0/5

In vitro activation of TCR $V\beta 8^+$ T-cell line TLMSB-L9 by SEB leads to transfer of EAE *in vivo* (a). Cells of TCR $V\beta 8^+$ encephalitogenic T-cell line TLMSB-L9 specific for MBP peptide Ac1-11 were activated either with $0.001 \mu\text{g ml}^{-1}$ SEB or $10 \mu\text{g ml}^{-1}$ Ac1-11 in the presence of syngeneic antigen-presenting cells or kept inactivated as shown. For disease induction, 5×10^6 T cells were transferred 3 days after the stimulation. The number of diseased mice was recorded daily as described for Fig. 1. As controls (b), non-encephalitogenic MBP peptide Ac1-11-specific T-cell clone TCMSB-C1 and normal splenocytes were activated and transferred in the same way as described for TLMSB-L9. Three days after stimulation, cells were directly double-stained according to standard procedures¹² with fluorescein-conjugated anti-IL-2-receptor antibody and phycoerythrin-conjugated anti-mouse CD4 antibody (all from Pharmingen, CA). Stained cells were analysed as described¹² and data are presented as the ratio of mean-fluorescence intensity of IL-2R stains of activated versus unactivated cells. ND, not done.

cells *in vitro* with superantigen. The encephalitogenic line, TLMSB-L9, was stimulated for 3 days with SEB or Ac1-11 and transferred into naive mice. SEB can serve as well as nominal bantigen in transforming cells from an inactive, non-encephalitogenic state to an encephalitogenic state (Table 2). SEB stimulation activates $V\beta 8^+$ T cells, indicated by the increase of interleukin-2 receptor (IL-2R) surface expression (Table 2). It is not possible to induce encephalitogenicity in the TCR $V\beta 8^+$ non-encephalitogenic T-cell clone TCMSB-C1, or in plain splenocytes by *in vitro* stimulation with SEB (Table 2). Taken together, these results show that SEB can directly activate TCR $V\beta 8^+$ encephalitogenic cells. Because cells were activated *in vitro*, a systemic effect of SEB could not account for our observations.

Kappler and Marrack suggested that superantigens might play a role in the activation of autoreactive T cells in the induction of autoimmunity^{21,22}. Here we demonstrate that microbial superantigens can influence the course of autoimmune disease through direct action on pathogenic T cells, once autoreactivity has been initiated.

Note added in proof: Similar observations have been reported by J. M. Soos *et al.* *J. Immun.* **150**, 192A (1993). □

Received 24 June; accepted 2 August 1993.

- Allen, I. in *McAlpine's Multiple Sclerosis* (eds Matthews, W. B., Compston, A. & Allen I. V.) (Churchill Livingstone, Edinburgh, 1991).
- Poser, C. M. *J. Neurol. Sci.* **107**, 127-140 (1992).
- Alvord, E. C. J. *Experimental Allergic Encephalomyelitis: A Useful Model for Multiple Sclerosis* (ed. Alvord, E. C. J.) (Liss, New York, 1984).
- Oksenberg, J. *et al.* *Nature* **362**, 68-70 (1993).
- Zamvil, S. S. & Steinman, L. *A. Rev. Immun.* **8**, 579-621 (1990).
- Zamvil, S. S. *et al.* *J. exp. Med.* **167**, 1586-1596 (1988).
- Acha-Orbea, H. *et al.* *Cell* **54**, 263-273 (1988).
- Urban, J. L. *et al.* *Cell* **54**, 577-592 (1988).
- Zamvil, S. S. *et al.* *Nature* **324**, 258-260 (1986).
- Bell, R. *et al.* *J. Immun.* **150**, 4085-4092 (1993).
- Lehmann, P. V., Forsthuber, T., Miller, A. & Sercarz, E. *Nature* **358**, 155-157 (1992).
- Gaur, A., Fathman, C. G., Steinman, L. & Brocke, S. J. *Immun.* **150**, 3062-3069 (1993).
- Behlke, M. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 767-771 (1986).
- Soos, J. M., Schiftenbauer, J. & Johnson, H. M. *J. Neuroimmun.* **43**, 39-43 (1993).
- Rott, O., Wekerle, H. & Fleischner, B. *Int. Immun.* **4**, 347-353 (1992).
- White, J. *et al.* *Cell* **56**, 27-35 (1989).
- Kawabe, Y. & Ochi, A. *J. exp. Med.* **172**, 1065-1070 (1990).

- Rellahan, B. L. *et al.* *J. exp. Med.* **172**, 1091-1100 (1990).
- Kawabe, Y. & Ochi, A. *Nature* **349**, 245-248 (1991).
- Powell, M. B. *et al.* *Int. Immun.* **2**, 539-544 (1990).
- Marrack, P. & Kappler, J. *Science* **248**, 705-711 (1990).
- Paliard, X. *et al.* *Science* **253**, 325-329 (1991).
- Brocke, S. & Hahn, H. *Inf. Immun.* **59**, 4531-4539 (1991).

ACKNOWLEDGEMENTS. We thank J. Abrams, DNAX Research Institute, for purified monoclonal anti-TNF and control antibodies. This work was supported by grants from the NIH to L.S. and C.G.F. S.B. was a fellow of the Myasthenia Gravis Society, K.G. was supported by the Belgian National Fund for Scientific Research (NFWO). A.G. was supported by the ADA.

The *daf-4* gene encodes a bone morphogenetic protein receptor controlling *C. elegans* dauer larva development

Miguel Estevez*, Lilliana Attisano†, Jeffrey L. Wrana†, Patrice S. Albert*, Joan Massagué† & Donald L. Riddle*

* Molecular Biology Program and Division of Biological Sciences, 311 Tucker Hall, University of Missouri, Columbia, Missouri 65211, USA

† Memorial Sloan-Kettering Cancer Center, Howard Hughes Medical Institute, Cell Biology and Genetics Program, 1275 York Avenue, New York, New York 10021, USA

THE bone morphogenetic protein (BMP) family is a conserved group of signalling molecules within the transforming growth factor- β (TGF- β) superfamily^{1,2}. This group, including the *Drosophila decapentaplegic* (*dpp*) protein and the mammalian BMPs, mediates cellular interactions and tissue differentiation during development^{3,4}. Here we show that a homologue of human BMPs controls a developmental switch in the life cycle of the free-living soil nematode *Caenorhabditis elegans*. Starvation and overcrowding induce *C. elegans* to form a developmentally arrested, third-stage dauer larva⁵. The *daf-4* gene, which acts to inhibit dauer larva formation and promote growth, encodes a receptor protein kinase similar to the *daf-1*, activin and TGF- β receptor serine/threonine kinases. When expressed in monkey COS cells, the *daf-4* receptor binds human BMP-2 and BMP-4. The *daf-4* receptor is the first to be identified for any growth factor in the BMP family.

The *daf-4* gene was tagged by insertion of the Tc1 transposon. Mutations in the *daf-4* gene confer a temperature-sensitive (ts) dauer-constitutive phenotype, in which dauer larvae are formed in abundant food⁶. In addition, *daf-4* adults are small, reaching about two-thirds the length of the wild type, and they are partially defective in egg-laying (*Egl*)⁷. Two independent transposon-induced *daf-4* alleles, (*m490* and *m500*), and two spontaneous revertants of each were obtained from a strain that exhibits a high frequency of Tc1 transposition in the germ line⁸. Fragments of genomic DNA containing the Tc1 element inserted in *m490* and *m500* were cloned by methods described previously⁸. Southern blots of restriction-enzyme-digested genomic DNA from the two *daf-4* mutants and their respective revertants probed with genomic sequences flanking the transposon revealed a single hybridizing band 1.6 kilobases (kb) larger in the mutants than in the revertants (Fig. 1). This difference corresponded to the insertion of a 1.6-kb Tc1 element in the two *daf-4* mutants and excision of the element in the two revertants.

The sequences flanking the *m500::Tc1* insertion site were used as a probe to isolate four overlapping genomic clones, and three overlapping cDNA clones containing inserts of 2.9 kb (clone DR#191), 1.6 kb and 1.1 kb. Genomic clone DR#180 was positioned on the physical map of chromosome III (ref. 9) between the cloned genes *ced-4* and *mec-14*, which flank *daf-4*.

TABLE 1 Characterization of two transgenic lines carrying a *daf-4* cDNA heat-shock expression construct*

Strain	Per cent L1 larvae growing to adult†		Per cent dauer larvae recovering‡	
	No heat shock	Heat shock	No heat shock	Heat shock
DR1418§	ND	ND	0 (0/114)	11 (13/114)
DR1419	21 (49/236)	55 156/284)	3 (6/177)	21 (80/390)
<i>daf-4(e1364)</i>	0 (0/297)	0 (0/233)	0 (0/282)	0 (0/996)

ND, not determined.

* The heat-shock expression plasmid DR#216, was constructed by insertion of the *daf-4* cDNA from DR191 into the *KpnI* and *EagI* sites of the vector pPD49.78, a heat-shock expression vector containing the *hsp16-2* promoter²⁴ upstream of a polylinker. To ensure better levels of expression from the cDNA sequence, the *SacI/EagI* fragment from the cDNA was replaced with its genomic equivalent, which contains a 466-bp intron.

† Heat shock for 1 h at 31 °C was imposed on populations of larvae that had been synchronized at hatching²⁵.

‡ All progeny were tested for heat-inducible recovery after remaining as dauer larvae for 5 d at 25 °C.

§ DR1418 had fewer than five progeny per hermaphrodite.

|| Of 285 adults obtained from asynchronous populations of DR1419 without heat shock, 31% were wild-type in length and not egg-laying-defective, and the remainder were small and slightly Egl.

on the genetic map.

Microinjection-mediated marker rescue experiments¹⁰ were done to identify a genomic clone carrying a complete copy of *daf-4*. Twenty *daf-4(m72)* hermaphrodites were co-injected with genomic clone DR#180 and DNA carrying the dominant morphological marker, *rol-6(su1006)*, which confers a dominant roller phenotype¹⁰. Of 13 rescued F₁ hermaphrodites, one continued to produce non-dauer progeny after the first generation. It was used to establish a permanent transgenic line, DR1269. A second transgenic line, DR1378 (not carrying the *rol-6* marker), was obtained from a similar experiment. Both lines produced non-Egl, wild-type length adult progeny, small adult progeny, and 20% dauer progeny, indicating that the *daf-4* transgene was

maintained as an unstable extrachromosomal array¹⁰. Thirteen control *daf-4* hermaphrodites (injected with *rol-6* and cosmid ZK372 from the chromosome III region adjacent to *daf-4*) produced 45 dauer larva progeny with a roller phenotype. The production of roller progeny confirmed that the microinjection was successful, but the non-*daf-4* cosmid failed to permit non-dauer development at 25 °C.

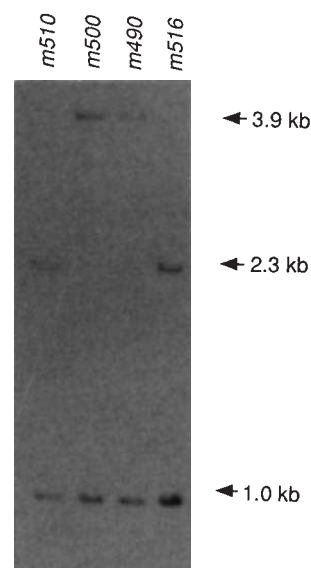
Wild-type *daf-4* activity is required both for preventing entry into the dauer stage and for promoting exit from the dauer stage⁵. The 2.9-kb cDNA insert detected two messages, 2.9 kb and 2.0 kb in size, on a northern blot of poly(A)⁺ mRNA from eggs, larvae and adults (manuscript in preparation). To determine whether the 2.9-kb *daf-4* cDNA can convey both wild-type functions, twenty hermaphrodites of genotype *daf-4(e1364)* were injected with DR#216 plasmid DNA (Table 1) and kept at 25 °C for three days to select for transgenic F₁ progeny. Three progeny grew to adulthood. Their dauer larva siblings were heat-shocked at 31 °C for 1 h, then returned to 25 °C to test for heat-inducible *daf-4* activity. Two days later, fourteen of these dauer larvae had resumed development. We obtained two independent transgenic lines (DR1418 and DR1419) that continued to show heat-inducible recovery of their dauer larva progeny.

Heat-shock of L1 larvae induced non-dauer development in DR1419 (Table 1). Some animals from DR1419 underwent non-dauer development even without heat-shock, presumably because expression of the transgene at 25 °C was sufficient in some animals to prevent dauer larva development. Control *daf-4* hermaphrodites that were either not injected, or were injected with the parent expression vector, produced only dauer larva progeny at 25 °C. These could not be induced to resume development by heat shock. Expression of *daf-4* either in the L1 or the dauer stage confers wild-type function sufficient to rescue the dauer-constitutive phenotype (Table 1).

Both strands of the DR#191 cDNA clone, and at least one strand of each of the other two cDNA clones, were sequenced (Fig. 2a). The predicted protein shows a pattern of hydrophobic-

FIG. 1 Detection of the *daf-4* gene. Genomic DNA from the mutants *daf-4(m490::Tc1)*; *daf-12(m20)* and *daf-4(m500::Tc1)*; *daf-12(m20)*, and their respective revertants, *daf-4(m516)* and *daf-4(m510)*, was digested with *EcoRI*, electrophoresed through a 1% agarose gel and blotted onto nitrocellulose. This blot was probed with a 0.7-kb sequence flanking the transposon insertion in *daf-4(m490::Tc1)*. The hybridizing band is 1.6 kb larger in the mutant lanes (labelled *m490* and *m500*) than in the revertant lanes (labelled *m510* and *m516*). The 1.0-kb band that does not shift revealed that the probe spans an *EcoRI* site. This was later confirmed by sequencing.

METHODS. Transposon mutagenesis: populations of the mutator strain RW7097 were grown at 20 °C, quantified and screened visually for the presence of constitutively formed dauer larvae as described previously⁸. Two alleles of *daf-4* were found. Both were recessive and temperature-sensitive (ts), with all animals arresting development at the dauer stage at 25 °C, and about half arresting at the dauer stage at 15 °C. Previous analysis of amber nonsense alleles of two dauer-constitutive genes, *daf-4* and *daf-7*, suggested that the phenotype of null alleles is ts⁶. A screen for revertants of *daf-4(m490)* used *daf-4(m490)*; *mut-6(st702)* as the parent strain. Cultures were established on agar plates at 15 °C, and maintained until wild-type length adults were observed. A spontaneous revertant of *daf-4(m500)* arose in the stock of *daf-4(m500)*; *mut-6(st702)* before a screen was initiated. The initial *daf-4* isolates and one revertant of each were crossed into Bristol genetic backgrounds to produce genetically stable strains carrying fewer copies of Tc1 (ref. 8) for the purpose of genomic DNA preparation. The *daf-12 X* mutation²⁶ was used to suppress the dauer-constitutive phenotype and permit more efficient growth of cultures. Cloning: genomic DNA was prepared from nematodes grown in liquid culture, but the freezing step before the proteinase K digestion was omitted²⁵. Genomic DNA from the mutants and revertants was digested with either *EcoRI/BamHI* or *XbaI/SstI*, and resolved on a 1% agarose gel. The gel was blotted onto nitrocellulose (Micron Separations) and probed with Tc1 to detect novel Tc1 containing fragments⁸. A novel 1.9-kb *EcoRI/BamHI* fragment and a novel 2.3-kb *XbaI/SstI* fragment were present in the mutant DNAs. To



clone these fragments, an *EcoRI/BamHI* and *XbaI/SstI* digest of mutant genomic DNA was resolved on a 1% agarose gel, the DNA was extracted from the gel slices by electroelution, and the fragments subcloned into the pBluescript KS(+) vector (Stratagene, La Jolla, CA). Two positive recombinant plasmids were isolated: DR#211 contains the 2.3-kb *XbaI/SstI* fragment from *daf-4(m490::Tc1)*, and DR#212 contains the 1.9-kb *EcoRI/BamHI* fragment from *daf-4(m500::Tc1)*. The transposon sequence in the DR#211 plasmid was removed by digestion with *EcoRV*, which cuts in the terminal repeats of Tc1, and the remaining flanking sequence was used as a probe.

ity characteristic of a transmembrane receptor with extracellular and cytoplasmic domains. The N-terminal hydrophobic sequence has the characteristics of a signal peptide with a putative cleavage site (AVT/E) after amino acid 31 (Fig. 2a). This is followed by a cysteine-rich domain, and a second 29-amino-acid hydrophobic sequence which putatively spans the cell mem-

brane. There are four potential glycosylation sites in the putative 196-amino-acid extracellular domain.

The predicted cytoplasmic portion of the *daf-4* protein contains a protein kinase domain related to those found in *daf-1* (ref. 8) and the type II receptors of TGF- β (ref. 11) and activin^{12,13} (Fig. 2b). These receptors all have a serine/threonine-

a

```

-128      TTTTTTTTTT
ATTAAATAAGTTTGCATCTATTTTCTCAATAAAACATTGTGATCTTAATTATTTATTT
ATTTTACTCTCAACTTTAAAAAAACGTTACAATAATCAACTACCTTCAATTTTCAG

+1  ATG AAT CAG AAA GGG ACA GTA CGT CTC AAG GCA CTT GTG CTC ATT
1   Met Asn Gln Lys Gly Thr Val Arg Leu Lys Ala Leu Val Leu Ile

46  TGC CTT CCA CTC TTT TTA ATT GCC ACA CCG GTT CCT GTT GCA GTC
16  Cys Leu Pro Leu Phe Leu Ile Ala Thr Pro Val Pro Val Ala Val

91  ACG GAA GAC GAT CGT CAG GAC CGA ATC GAA TCG GAA GCC GCC GAG
31  Thr Glu Asp Asp Arg Gln Asp Arg Ile Glu Ser Glu Ala Ala Glu

136 AAA GAG TGG GCT AAC ACC CTA GTC AGC AAA GTA GCA CAG TCA AAT
46  Lys Glu Trp Ala Asn Thr Leu Val Ser Lys Val Ala Gln Ser Asn

181 GGA ACT GGC ACT ATT AAG GTG TCG GCA CCA GCG AAG CCC ACG TTA
61  Gly Thr Gly Thr Ile Lys Val Ser Ala Pro Ala Lys Pro Thr Leu

226 AGG AGA ATG GAC AAT GAG GAG GAT GAA GTG ATT TCG ATT GAA TGT
76  Arg Arg Met Asp Asn Glu Glu Asp Glu Val Ile Ser Ile Glu Cys

271 GTA TAC TAC GAC GAG ATG GAA TGC GAA AAG AGT GGC GAT TGT GAA
91  Val Tyr Tyr Asp Glu Met Glu Cys Glu Lys Ser Gly Asp Cys Glu

316 ATC ACA AAG AAG ACG TGT TAC AGT GAA GCT CAT CTG AAA GCC GTC
106 Ile Thr Lys Lys Thr Cys Tyr Ser Glu Ala His Leu Lys Ala Val

361 GGA TGC TTG GCC GTA TTT GGG CTG CCC ACG CAA GAG ATA AAC TCT
121 Gly Cys Leu Ala Val Phe Gly Leu Pro Thr Gln Glu Ile Asn Ser

406 ACA GAA CCA TAC CTG AAA GTT GAT AAA CCG CAA TAC AAA AGT CTG
136 Thr Glu Pro Tyr Leu Lys Val Asp Lys Pro Gln Tyr Lys Ser Leu

451 GGA TGT ATG CCG TAT CAG CAC GCG GAC TCG ATG AAC TGC GAG AAC
151 Gly Cys Met Pro Tyr Gln His Ala Asp Ser Met Asn Cys Glu Asn

496 GAA AGC TCC TGT CGC CAA GGA CGA TCA TTT CGC GGC GGA ATT GGA
166 Glu Ser Ser Cys Arg Gln Gly Arg Ser Phe Arg Gly Gly Ile Gly

541 ATG TGC TGC TGT TCG ACC AAC AAT TGC AAC ATG CCG GAT CTC ATC
181 Met Cys Cys Cys Ser Thr Asn Asn Cys Asn Met Pro Asp Leu Ile

586 GAG ATG GTC AAC CCG TCA TTG AAG AAG GAT TCG GAT AAT TCT GCA
196 Glu Met Val Asn Pro Ser Leu Lys Lys Asp Ser Asp Asn Ser Cys

631 CTT CTG TGG GCT TCA ACT CCA TCC AAC ATG GAT TTG GAA TCT CTA
211 Leu Leu Trp Ala Ser Thr Pro Ser Asn Met Asp Leu Glu Ser Leu

676 GAC AAG TTC CCA TTC TAC TGG ATC ATC ATA ATC GCA CTG TCA GTG
226 Asp Lys Phe Pro Phe Tyr Trp Ile Ile Ile Ile Ala Leu Ser Val

721 ATC CTT TGT ATT GCA CTC TTG ATC TTG GCA TAC GTT GGT TGS AAA
241 Ile Leu Cys Ile Ala Leu Leu Ile Leu Ala Tyr Val Gly Trp Lys

766 TTT CAG CAG AAT AAG AAG GAG GAG ATA AAA AAG CAG CAG AAA ATT
256 Phe Gln Gln Asn Lys Lys Glu Glu Ile Lys Lys Gln Gln Lys Ile

811 AAA TTT GAC ATG GAA AAA ACT GAC GCA CTG GAA GCT GGA AAT GTG
271 Lys Phe Asp Met Glu Lys Thr Asp Ala Leu Glu Ala Gly Asn Val

856 CCT CTC GTG GAG CCG GAA GAG GAA ATG ATT GAA ATG GTG GAG ACA
286 Pro Leu Val Glu Pro Glu Glu Glu Met Ile Glu Met Val Glu Thr

901 CCA AAG GAG CTG CCG ATC ACT GAT TTT CAG TTA ATC TCA AAG GGA
301 Pro Lys Glu Leu Pro Ile Thr Asp Phe Gln Leu Ile Ser Lys Gly

946 CGT TTT GGA AAA GTT TTC AAG GCG CAA TAC ACC CCA GAC AGT GGT
316 Arg Phe Gly Lys Val Phe Lys Ala Gln Tyr Thr Pro Asp Ser Gly

991 GAG AAA CGA TTG GTT GCG GTG AAG AAA TTG AAC GAA TTC CAA AAA
331 Glu Lys Arg Leu Val Ala Val Lys Lys Leu Asn Glu Phe Gln Lys

1036 GCG AGC TTT TTG GCG GAG AAA AGA ATC TTT GAT GAG CTT AAC GAG
346 Ala Ser Phe Leu Ala Glu Lys Arg Ile Phe Asp Glu Leu Asn Glu

1081 TAC CCG AAG TGG TAT AAA AGT ATC GTC GAA TTT GTG TGC CCC GAG
361 Tyr Pro Lys Trp Tyr Lys Ser Ile Val Glu Phe Val Cys Pro Glu

1126 AAA ATT GGT GAT GAG TAT TGG ATT GTG ACC GAA TTT CAT GAG CGA
376 Lys Ile Gly Asp Glu Tyr Trp Ile Val Thr Glu Phe His Glu Arg

1171 CTT AGT CTC TAT GAG TTA CTC AAG AAT AAT GTA ATT AGT ATT ACA
391 Leu Ser Leu Tyr Glu Leu Leu Lys Asn Asn Val Ile Ser Ile Thr

```

```

1216 TCT GCC AAT CGA ATT ATT ATG TCA ATG ATC GAT GGG CTT CAA TTC
406 Ser Ala Asn Arg Ile Ile Met Ser Met Ile Asp Gly Leu Gln Phe

1261 TTG CAT GAC GAC AGG CCG TAC TTC TTT GGA CAC CCA AAG AAG CCA
421 Leu His Asp Asp Arg Pro Tyr Phe Phe Gly His Pro Lys Lys Ser

1306 ATT ATT CAC AGA GAT ATC AAG TCG AAG AAC ATT CTT GTG AAA AGC
436 Ile Ile His Arg Asp Ile Lys Ser Lys Asn Ile Leu Val Lys Ser

1351 GAC ATG ACC ACA TGC ATC GCG GAC TTT GGC CTC GCC CGA ATC TAC
451 Asp Met Thr Thr Cys Ile Ala Asp Phe Gly Leu Ala Arg Ile Tyr

1396 AGC TAT GAC ATT GAG CAA AGC GAT TTG CTG GGT CAA GTG GGA ACG
466 Ser Tyr Asp Ile Glu Gln Ser Asp Leu Leu Gly Gln Val Gly Thr

1441 AAA CGC TAC ATG TCA CCA GAG ATG CTC GAA GGA GCA ACC GAA TTC
481 Lys Arg Tyr Met Ser Pro Glu Met Leu Glu Gly Ala Thr Glu Phe

1486 ACG CCA ACT GCT TTC AAA GCG ATG GAC GTT TAT TCG ATG GGA CTC
496 Thr Pro Thr Ala Phe Lys Ala Met Asp Val Tyr Ser Met Gly Leu

1531 GTC ATG TGG GAA GTC ATA TCC CGT ACC AAG TTG CAC CAA ACC GAC
511 Val Met Trp Glu Val Ile Ser Arg Thr Lys Leu His Gln Thr Asp

1576 GAG CCA CCA AAC TAC CAA ATG CCG TTT CAA GTC ATC GGA TTT GAT
526 Glu Pro Pro Asn Tyr Gln Met Pro Phe Gln Val Ile Gly Phe Asp

1621 CCC ACA ATT GGT CGC ATG AGA AAT TAT GTG GTG AGC AAG AAG GAA
541 Pro Thr Ile Gly Arg Met Arg Asn Tyr Val Val Ser Lys Lys Glu

1666 CGA CCA CAA TGG AGA GAT GAG ATT ATT AAA CAT GAA TAT ATG AGT
556 Arg Pro Gln Trp Arg Asp Glu Ile Lys His Glu Tyr Met Ser

1711 CTG CTA AAA AAG GTT ACG GAA GAG ATG TGG GAT CCC GAA CCG TGT
571 Leu Leu Lys Lys Val Thr Glu Glu Met Trp Asp Pro Glu Ala Cys

1756 GCA CGG ATT ACA GCT GGA TGT GCG TTC GCG AGG GTA TGG AAT CAT
586 Ala Arg Ile Thr Ala Gly Cys Ala Phe Ala Arg Val Trp Asn His

1801 ATT ATG AGC TCA CCG GAC AGT TCG GAG GGG TAT CAC AGT GGA AGT
601 Ile Met Ser Ser Pro Asp Ser Ser Glu Gly Tyr His Ser Gly Ser

1846 AGT ATG AAG AAT CCG GGT GTT GAT GAT GTT GAA CAA AGT GAG AAG
616 Ser Met Lys Asn Arg Gly Val Asp Asp Val Glu Gln Ser Glu Lys

1891 CCG GAG GGT ATC GAA GAA ATG CAA CAT TAT CAT GCG TCG TCA CCC
631 Pro Glu Gly Ile Glu Glu Met Gln His Tyr His Ala Ser Ser Pro

1936 AGC AAG AGA CAA CAT CCT AGT CCG AAT CCG TTT TTT GAT TCG TGT
646 Ser Lys Arg Gln His Pro Ser Pro Asn Pro Phe Phe Asp Ser Cys

1981 CCA CCA CCT CCG CCG ATT CCT GTG ATC CTT GAA AAT GGA GGA ATC
661 Pro Pro Pro Pro Pro Ile Pro Val Ile Leu Glu Asn Gly Gly Ile

2026 CTC CAA CCA GAC AAC GCG GAG CCA GAG CCC GAA GAG CTG CCG GAT
676 Leu Gln Pro Asp Asn Ala Glu Pro Glu Pro Glu Glu Leu Pro Asp

2071 CTC CCG ATA GTC GAG AAA ATC TAC GAT ATT GCG ACG AAC ATG CTA
691 Leu Pro Ile Val Glu Lys Ile Tyr Asp Ile Ala Thr Asn Met Leu

2116 TTC TCA AGA GAA GAG CTT GAC CTT ATG AAT GCT CAA CGG CAA GTT
706 Phe Ser Arg Glu Glu Leu Asp Leu Met Asn Ala Gln Arg Gln Val

2161 GAA TAC GAA GCT GGG GCA GAC ACT CGA GCG TCG ACA CCA ACG CCT
721 Glu Tyr Glu Ala Gly Ala Asp Thr Arg Ala Ser Thr Pro Thr Pro

2206 TCA GGG ACG TTT GGA ACG TTT ACA ACA TAA
736 Ser Gly Thr Phe Gly Thr Phe Thr Thr

2236 CTATTCATTCAAATCTAACATTATTTATATTATTACTCCCATCTCTCTCTATT
CTAGTCTATATGTAGCGGTGGACACCTCAGCGGCCCTCGCCGCTGGCCACCACTTG
TCTCTACTTACGCGTCCCTAATGATCTCAATTGAGCACAACCGGTGTATATCAAC
TTACAGACGTTCAATATAAGAACCGCTAAAGTTCTATGCTCTCGTCTAAACCTCT
TCCAAAATCTCTGATCTCTTGCACTGTCTCTCACTCAACCAAGTTGCCCACTCTCATCC
AACTAACTCAAAAACCTTACCTCAGCCATGGGGCATCCCATATTCTCTCATTCATTCAC
CTATATTCATTGATCTGTGATACTTTGATTGGCAGTTTACATCACACAAGAACTCTTT
TTTGTATATCAGTTGATTATTTATCTTCATTTTACACTTAAATTAATATGTT
TCTT

```


Two independently isolated mutants, *m490* and *m500*, carry TcI insertions at the same site in the region encoding kinase subdomain XI (Fig. 2a). The insertion site was determined for both mutants by sequencing the flanking regions. A previously described ethylmethane sulphonate-induced, amber-suppressible

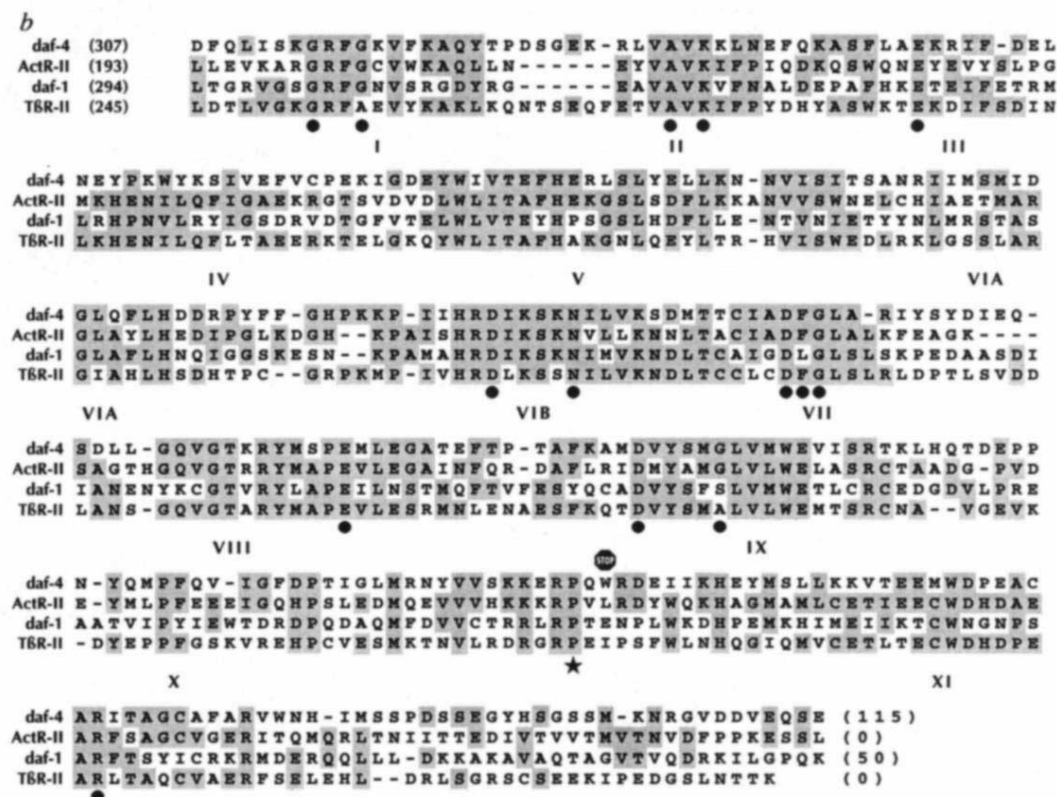


FIG. 2 Sequence analysis. a, The *daf-4* cDNA sequence and translated amino-acid sequence. The nucleotide sequence determined from the *daf-4* cDNAs (DR#191, DR#192, and DR#193) and DR#180 genomic DNA is shown. The smaller cDNA clones overlap completely with the DR#191 cDNA, except that DR#191 ends before the polyadenylation signal. In addition, about two-thirds of the cDNA sequence was confirmed by genomic sequence from DR#180, including both the 5' and 3' untranslated regions (UTR). In DR#191 there are 128 bp upstream of the putative initiator methionine codon, numbered 1 in the translated amino-acid sequence. The *daf-4* message is *trans*-spliced with the 22-nucleotide SL1 leader sequence²⁷. A signal used for *trans*-splicing is underlined in bold. The longest open reading frame extends for 2,232 bp and translates into a 744-amino-acid, 86K protein with a presumed signal peptide domain (double underline), glycosylation sites (underline), transmembrane domain (double underline), and protein kinase domain (in brackets) beginning at the nucleotide binding site (SKGRFG) in subdomain I. The first amino acid at this site is predicted to be a serine instead of glycine. Similar conservative substitutions have been found in other kinases^{12,15}. A cysteine-rich extracellular motif that is similar in all *daf*, activin and TGF- β receptor sequences is boxed. An acceptable polyadenylation signal (bold underline) is near the 3' end of the 476-bp 3' UTR. An upstream polyadenylation signal (bold overline) is in a position consistent with the production of a 2.0-kb mRNA. An arrowhead indicates the position at which Tc1 elements inserted in mutants *m490* and *m500*. A filled circle is over the site of the GC to

METHODS. Isolation of cDNA clones: a lambda ZAP cDNA library²⁸ was probed with the sequences flanking the Tc1 insertion site in *m500::Tc1*. Three cDNA clones were identified (DR#191, DR#192, and DR#193), and the internal pBluescript SK (–) plasmid was excised using protocols supplied by Stratagene. The 5' end of the *daf-4* transcript was PCR-amplified from egg mRNA using the SL1 *trans*-spliced leader sequence and an internal *daf-4* primer. DNA was sequenced by the dideoxy chain termination method using Sequenase II (US Biochemical). The *daf-4* (*m72*) mutation⁶ was localized by chemical mismatch detection¹⁶, and then sequenced using the PCR sequencing technique (Perkin-Elmer-Cetus).

mutation⁶, *daf-4(m72)*, was located within the coding region by chemical mismatch detection¹⁶. Direct sequencing of *daf-4(m72)* DNA amplified by the polymerase chain reaction (PCR) revealed a GC to AT transition corresponding to cDNA position 1,676 (Fig. 2a, b). This mutation changes a tryptophan codon to an amber stop codon in the coding sequence for kinase subdomain X. The sites of these mutations suggest that an intact kinase domain is required for normal *daf-4* function.

We tested the ability of the *daf-4* receptor to interact with representative members from each of the major subgroups of the TGF- β superfamily, namely the TGF- β family, the BMP (or DVR, for *decapentaplegic*-Vg-related) family¹ and the activin/inhibin family. Monkey COS-1 cells were transfected with either *daf-4* cDNA or other related receptor sequences (T β R-II, *daf-1* and ActR-IIB) inserted into the mammalian expression vector pCMV5. Cell layers were incubated with ¹²⁵I-labelled TGF- β 1, ¹²⁵I-activin A or ¹²⁵I-BMP-2, affinity-crosslinked, and analysed by SDS-PAGE. Of the four receptors tested, only T β R-II formed a complex with ¹²⁵I-TGF- β , and only ActR-IIB formed a complex with ¹²⁵I-activin A (data not shown). The *daf-1* receptor did not interact with any of the ligands tested. By contrast, COS-1 cells expressing *daf-4* yielded an affinity-labelled complex of *M_r* 100K when cross-linked to ¹²⁵I-BMP-2 (Fig. 3a).

For more detailed analysis of ligand specificity, the *daf-4* receptor was modified to improve the signal intensity and to minimize background. The HA1 epitope of influenza virus haemagglutinin was incorporated into the 3' end of the receptor, allowing immunoprecipitation of the affinity-labelled receptor. In addition, the *C. elegans* start site was replaced with a mam-

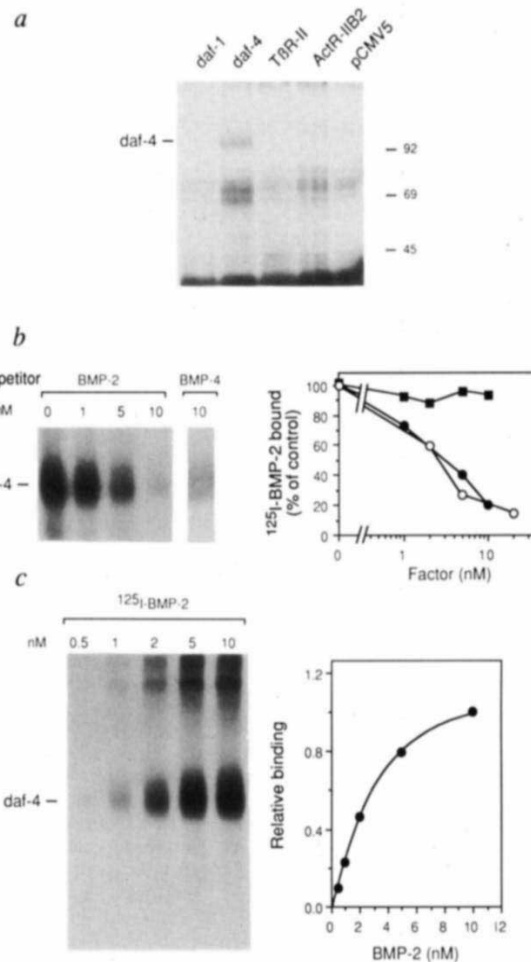
malian consensus start site. Transfected cells were incubated with ¹²⁵I-BMP-2 in the presence of excess unlabelled factors, and the affinity-labelled receptor was immunoprecipitated. The binding of ¹²⁵I-BMP-2 was competed by unlabelled BMP-2 or the highly related factor BMP-4 with similar potency (dose for 50% inhibition ID₅₀ = 2–5 nM), but not by TGF- β 1 (Fig. 3b). Half-maximal saturation of *daf-4* receptor was obtained with 2–5 nM ¹²⁵I-BMP-2 (Fig. 3c), which is in the range of biologically active BMP-2 concentrations in various mammalian cell systems^{4,17}. These results demonstrate that the *daf-4* receptor is able to bind specifically and with high affinity to mammalian BMP-2 and BMP-4. The two human factors are about 92% identical to each other¹.

BMP binding strongly suggests that the natural ligand for the *daf-4* receptor is a *C. elegans* BMP homologue, but comparison of the receptor sequence with that of a BMP receptor is not possible as no such receptor has yet been found. Nevertheless, similarities of the *daf-1* and *daf-4* receptors to the activin and TGF- β receptors, in both the cytoplasmic and the extracellular domains, suggest that these receptors have similar functions and may bind similar ligands.

Although the *daf-1* receptor produced negative results in the COS cell-binding assay, it is possible that the *daf-1* and *daf-4* proteins form a heterodimeric receptor in *C. elegans*. Alternatively, these receptors may function sequentially. Gene expression or immunolocalization studies should reveal whether *daf-1* and *daf-4* are expressed in the same or in different cells. It is also possible that *daf* function is required for development or maintenance of cells that transmit a signal for non-dauer devel-

FIG. 3 *daf-4* encodes a BMP receptor. a, Monkey COS-1 cells transiently transfected with pCMV5 expression vectors for either *daf-1* (ref. 8), *daf-4*, T β R-II (refs 11, 29), ActR-IIB2 (ref. 13) or pCMV5 alone were incubated with 250 pM ¹²⁵I-BMP-2 for 3 h at 4 °C and affinity-crosslinked using disuccinimidyl suberate as previously described for TGF- β 1 (ref. 30). Affinity-labelled samples were analysed by SDS-PAGE under reducing conditions followed by autoradiography. A specific BMP-2-labelled species of *M_r* 100K is present only in COS-1 cells expressing the *daf-4* receptor. b and c, Monkey COS-1 cells were transiently transfected with *daf-4* in pCMV5 containing a mammalian consensus start site and HA1 epitope of influenza virus haemagglutinin (HA) or with pCMV5 alone. b, Cells were incubated with 500 pM ¹²⁵I-BMP-2 and the indicated concentrations of unlabelled BMP-2 (filled symbols; gift from J. Wozney), BMP-4 (open symbols; gift from A. H. Reddi) or TGF- β 1 (squares; R&D Systems) and c, at the indicated concentrations of ¹²⁵I-BMP-2. Affinity-labelled cells were lysed and immunoprecipitated with monoclonal anti-HA1 antibody 12CA5 (BAbCO) and samples were analysed by SDS-PAGE followed by phosphorimage densitometry of the labelled bands.

METHODS. For initial COS-1 cell transient transfections, the *KpnI/PstI* fragment of *daf-4* cDNA clone DR#191 was subcloned into pCMV5 (ref. 13). For immunoprecipitations, the HA1 epitope of influenza virus haemagglutinin was introduced at the 3' end of the *daf-4* receptor (nucleotide 1,873) as described previously for T β R-II (ref. 29). The *C. elegans* *daf-4* translational start site CAGATGA was replaced with a mammalian consensus start site, ACCATGG, using PCR. For transient transfections, monkey COS-1 cells were incubated with 2 μ g ml⁻¹ plasmid diluted in Dulbecco's modified Eagle's medium (DMEM) containing 10% NuSerum (Collaborative Research), 400 μ g ml⁻¹ of DEAE-dextran and 100 μ M chloroquine³⁰; 24 h after transfection, the cells were trypsinized and reseeded into 6-well dishes and subsequently assayed for ligand binding 72 h post-transfection. Human recombinant BMP-2 was iodinated by the chloramine-T method (ref. 13). Affinity-labelled cells in 6-well dishes were solubilized in 0.5 ml lysis buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA) in the presence of protease inhibitors (same as for affinity labelling) at 4 °C for 20–30 min. Insoluble debris was removed by microcentrifugation at 10,000 r.p.m. for 5 min and HA-tagged receptors were immunoprecipitated by incubating the lysate with 2.5 μ g monoclonal antibody 12CA5 (BAbCO) for 1 h at 4 °C, followed by adsorption to protein A-Sepharose (Pharmacia). Beads were washed six times in lysis buffer with 0.1% Triton X-100 and bound protein was eluted by heating in SDS-PAGE sample buffer containing dithiothreitol.



opment. For example, ablation of a subset of amphid neurons with a laser microbeam results in a dauer-constitutive phenotype¹⁸. However, transient expression of the *daf-4* cDNA in the L1 or dauer stages is sufficient to rescue the mutant phenotype. We conclude that the cells required for non-dauer development are present in the mutants, and that rescue of the *daf-4* phenotype occurs without cell proliferation or differentiation.

The BMP/DVR proteins are a heterogeneous subgroup of related factors within the TGF- β superfamily^{1,2}. Many of the BMPs have been isolated on the basis of either their ability to induce ectopic bone formation or by their sequence similarity to related ligands. But the function of BMPs is not limited to their osteoinductive properties. They can also regulate neurogenic¹⁹ and myogenic differentiation¹⁷, and chick limb-bud development⁴. *Xenopus* versions of BMP-2, 4 and 7 have been implicated in the formation of ventral embryonic mesoderm^{20,21}. In *Drosophila*, the *dpp* gene is important for dorsal-ventral patterning in the embryo as well as imaginal disk formation^{22,23}. The finding that the *C. elegans daf-4* gene encodes a receptor for a BMP/*dpp*-like factor provides new evidence that signalling by this group of factors is fundamental in widely different developmental processes in metazoans. □

Received 13 April; accepted 26 July 1993.

1. Lyons, K. M., Jones, C. M. & Hogan, B. L. M. *Trends Genet.* **7**, 408–412 (1991).
2. Rosen, V. & Thies, R. S. *Trends Genet.* **8**, 97–102 (1991).
3. Kingsley, D. M. *et al. Cell* **71**, 399–410 (1992).
4. Niswander, L. & Martin, G. R. *Nature* **361**, 68–71 (1993).
5. Golden, J. W. & Riddle, D. L. *Dev. Biol.* **102**, 368–378 (1984).
6. Golden, J. W. & Riddle, D. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 819–823 (1984).
7. Trent, C., Tsung, N. & Horvitz, H. R. *Genetics* **104**, 619–647 (1983).

8. Georgi, L. L., Albert, P. S. & Riddle, D. L. *Cell* **61**, 635–645 (1990).
9. Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821–7825 (1986).
10. Meilo, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. *EMBO J.* **10**, 3959–3970 (1991).
11. Lin, H. Y., Wang, X.-F., Ng-Eaton, E., Weinberg, R. & Lodish, H. F. *Cell* **68**, 775–785 (1992). Erratum, *Cell* **70**, 1069 (1992).
12. Matthews, L. & Vale, W. *Cell* **65**, 973–982 (1991).
13. Attisano, L., Wrana, J., Cheifetz, S. & Massagué, J. *Cell* **68**, 97–108 (1992).
14. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. *Science* **259**, 1157–1161 (1993).
15. Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42–52 (1988).
16. Montandon, A. J., Green, P. M., Gianelli, F. & Bentley, D. R. *Nucleic Acids Res.* **17**, 3347–3358 (1989).
17. Yamaguchi, A. *et al. J. Cell Biol.* **113**, 681–687 (1991).
18. Bargmann, C. & Horvitz, H. R. *Science* **251**, 1243–1246 (1991).
19. Paralkar, V. M., Weeks, B. S., Yu, Y. M., Kleinman, H. K. & Reddi, A. H. *J. Cell Biol.* **119**, 1721–1728 (1992).
20. Dale, L., Howes, G., Price, B. M. J. & Smith, J. C. *Development* **115**, 573–585 (1992).
21. Jones, C. M., Lyons, K. M., Lapan, P. M., Wright, C. V. E. & Hogan, B. L. M. *Development* **115**, 639–647 (1992).
22. Irish, V. F. & Gelbart, W. M. *Genes Dev.* **1**, 868–879 (1987).
23. Panganiban, G. F., Rashka, K. E., Neitzel, M. D. & Hoffmann, F. M. *Molec. cell. Biol.* **10**, 2669–2677 (1990).
24. Jones, D., Russnak, R. H., Kay, R. J. & Candido, E. P. *J. biol. Chem.* **261**, 12006–12015 (1986).
25. Sulston, J. E. & Hodgkin, J. In *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 587–606 (Cold Spring Harbor Laboratory Press, New York, 1988).
26. Riddle, D. L., Swanson, M. M. & Albert, P. S. *Nature* **290**, 668–671 (1981).
27. Krause, M. & Hirsh, D. *Cell* **49**, 753–761 (1987).
28. Barstead, R. J. & Waterston, R. H. *J. biol. Chem.* **264**, 10177–10185 (1989).
29. Wrana, J. L. *et al. Cell* **71**, 1003–1014 (1992).
30. Massagué, J. *Meth. Enzym.* **146**, 174–195 (1987).

ACKNOWLEDGEMENTS. We thank A. Fire for pPD49.78 and for help with microinjection, J. Kramer for the pRF4 plasmid containing the dominant *rol-6* mutation, R. Barstead for the cDNA library, C. Link and W. B. Wood for the EMBL4 *C. elegans* genomic library, A. Coulson for mapping a *daf-4* genomic clone, E. Stringham and P. Candido for advice, P. L. Larsen for PCR amplification of SL1 trans-spliced *daf-4* mRNA, and L. Hatley for processing the manuscript. Some of the *C. elegans* strains used in this work were obtained from the *Caenorhabditis* Genetics Center, which is supported by the NIH National Center for Research Resources. This work was supported by DHHS grants to D.L.R. and to J.M. M.E. was the recipient of an NSF predoctoral fellowship, and L.A. and J.L.W. were recipients of postdoctoral fellowships from the Medical Research Council of Canada. The *daf-4* sequence has Genbank accession number L23110.

Somatic mutations in the thyrotropin receptor gene cause hyperfunctioning thyroid adenomas

Jasmine Parma*, Laurence Duprez*,
Jacqueline Van Sande*, Pascale Cochaux*†,
Christine Gervy*, Jean Mockel*,
Jacques Dumont* & Gilbert Vassart*†

* Institut de Recherche Interdisciplinaire and † Service de Génétique Médicale, Faculty of Medicine, University of Brussels, Campus Erasme, 808 route de Lennik, 1070 Brussels, Belgium

THE pituitary hormone thyrotropin stimulates the function, expression of differentiation and growth of thyrocytes by cyclic AMP-dependent mechanisms^{1–3}. Tissue hyperplasia and hyperthyroidism are therefore expected to result when activation of the adenyl cyclase–cAMP cascade is unregulated. This is observed in several situations^{4,5}, including when somatic mutations impair the GTPase activity of the G protein G_{sa} (refs 6, 7). Such a mechanism is probably responsible for the development of a minority of monoclonal hyperfunctioning thyroid adenomas^{6,8,9}. Here we identify somatic mutations in the carboxy-terminal portion of the third cytoplasmic loop of the thyrotropin receptor in three out of eleven hyperfunctioning thyroid adenomas. These mutations are restricted to tumour tissue and involve two different residues (aspartic acid at position 619 to glycine in two cases, and alanine at position 623 to isoleucine in one case). The mutant receptors confer constitutive activation of adenyl cyclase when tested by transfection in COS cells. This shows that G-protein-coupled receptors are susceptible to constitutive activation by spontaneous somatic mutations^{10,11} and may thus behave as proto-oncogenes.

Hyperfunctioning thyroid adenomas are discrete encapsulated neoplasms that are activated independently of thyroid-stimulating hormone (TSH), leading to hyperthyroidism and, as a result of the negative feedback exerted by thyroid hormones on TSH release, to silencing of thyroid function in extra-adenomatous tissue. Mutations have been found that activate the G protein G_{sa} in pituitary adenomas¹² and in some thyroid tumours^{6,8,9}. Mutation of specific residues within the carboxy-terminal portion of the third intracytoplasmic loop of the α_{1b} adrenergic receptor results in constitutive activation of phospholipase C^{10,11}, and this segment is probably critical for the activation of G proteins in the whole gene family. We therefore decided to look for somatic mutations in the third cytoplasmic loop of the TSH receptor in a series of eleven hyperfunctioning thyroid adenomas.

Each adenoma was identified *in vivo* by ¹³¹I scanning and, after its removal, by its encapsulation and high-iodide trapping *in vitro* compared with the quiescent surrounding tissue^{13,14}. The sensitivity of individual tumours to TSH was tested by comparing the accumulation of cAMP and inositol phosphates with extra-adenomatous tissue. In non-treated patients, adenomatous tissue is generally less sensitive to TSH¹⁴. cAMP concentrations are similar in slices of adenomatous and extra-adenomatous tissue incubated in the absence of TSH, but these are hard to determine because of differences in cell density and proportion of necrotic tissue.

Direct sequencing after polymerase chain reaction (PCR) amplification of the gene segment encoding residues 589 to 647 showed an identical T-to-C transition (non-coding strand) in two tumours (Fig. 1a, bottom), resulting in the replacement of Asp 619 by glycine (Fig. 2). The mutation created a *Kpn*I restriction enzyme digestion site and so could be demonstrated by Southern blotting of the two non-amplified DNA samples (results not shown). A third adenoma had an unusual mutation consisting of two contiguous base substitutions (C-G to T-A) (Fig. 1a, top). Sequencing of the cloned fragment demonstrated

TABLE 1 cAMP accumulation in cells carrying mutant TSH receptor constructs after treatment with TSH or forskolin

	TSH (mU ml ⁻¹)					Forskolin (10 ⁻⁵ M)
	0.00	0.01	0.10	1.00	10.00	
Transfected DNA pSVL	1.5 ± 0.1 (0.0)	ND	ND	ND	1.8 ± 0.1 (0.38)	44.3 ± 2.3
Wild-type TSH receptor	5.7 ± 0.1 (2.1)	7.5 ± 0.1 (2.9)	14.8 ± 0.4 (6.7)	19.1 ± 0.7 (8.6)	32.0 ± 3.4 (15.0)	69.0 ± 2.6
Asp 619 → Gly	13.8 ± 0.2 (7.0)	14.8 ± 0.6 (7.5)	13.9 ± 2.6 (6.5)	21.6 ± 0.2 (11.4)	30.5 ± 2.1 (16.4)	115.3 ± 5.7
Ala 623 → Ile	16.1 ± 0.8 (18.2)	18.4 ± 0.7 (21.2)	24.0 ± 0.7 (28.0)	24.9 ± 0.2 (28.2)	29.6 ± 1.3 (35.1)	85.0 ± 4.0

Transfected cells (1.5×10^5 cells seeded per dish) were incubated in Krebs-Ringer-HEPES (KRH) buffer containing Ro 20-1724 and Isobutylmethylxanthine (1 mM each) for 40 min, together with the concentrations of TSH or forskolin indicated. Results (means of triplicate transfections $\pm$ s.e.m.) are expressed as pmol intracellular cAMP per culture dish divided by the amount of DNA (in μ g) measured in parallel dishes. Results in parenthesis were obtained by subtracting from individual values the cAMP accumulated in cells transfected by the pSVL vector (1.5 pmol) and dividing by the human growth hormone accumulated in the culture medium to normalize for transfection efficiency. ND, not determined. pSVL-based constructs harbouring the two mutations were obtained by replacement in the wild-type construct²¹ of a CvnI/BstII segment (coordinates 1,620–2,185) with the corresponding segments from the mutant genes. These were obtained by PCR amplification of a 923-base-pair segment from tumour DNA using the primers 5'-GCTCTCCTGGGCAATGCTT-3' and 5'-ATGTTGTGGAGACCCTGCCT-3'. The sequences of the resulting mutated constructs (TSH receptor Asp 619 → Gly and Ala 623 → Ile, respectively) were verified by double-strand sequencing. COS-7 cells were grown in DMEM supplemented with 10% fetal bovine serum, penicillin 100 IU ml⁻¹, streptomycin 100 μ g ml⁻¹, fungizone 2.5 μ g ml⁻¹ and sodium pyruvate 1 mM. On day zero, cells were transfected with the various constructs (1.5 μ g DNA per dish) by the DEAE-dextran method²². To monitor the transfection efficiency, 0.5 μ g plasmid pXGH5 (Nichols Institute, Los Angeles) was added in each transfection. On day 2, cells were trypsinized and plated in 30-mm dishes (1.5×10^5 cells seeded per dish); 24 h later, medium was assayed for growth hormone (Pharmacia Diagnostics) and cells were preincubated in 1 ml KRH buffer for 40 min before incubation in the same medium supplemented with Ro 20-1724 and IBMX (10⁻³ M each) and TSH (Armour, Chicago) or forskolin. After 40 min incubation, medium was removed, cells were boiled in water and intracellular cAMP determined by radioimmunoassay²³. The accumulation of wild-type and mutated TSH receptor transcripts in transfected COS cells was analysed by northern blotting in parallel experiments. No significant difference could be detected (not shown). DNA for wild-type and mutant constructs was prepared in parallel to minimize differences in quality. Results were similar from two different DNA preparations and for different amounts of DNA constructs transfected per dish (1.0, 1.5 and 1.8 μ g).

that both substitutions were on the same allele (Fig. 1b), resulting in the replacement of Ala 623 by isoleucine (Fig. 2). The normal sequence was also detected in all cases, which is compatible with the mutation of a single allele but might also be the result of contamination of tumour tissue with normal cells (for example blood or fibroblasts). The extra-adenomatous tissue from the same patients showed only the normal sequence (Fig. 1a), which excludes polymorphism as an explanation for the

TABLE 2 Accumulation of ³H-inositol phosphates in cells carrying mutant TSH receptor constructs after treatment with TSH or ATP

	TSH (mU ml ⁻¹)		ATP (10 ⁻⁴ M)
	0	100	
Transfected DNA pSVL	5.4 ± 0.4 (0.00)	5.5 ± 0.1 (0.00)	22.1 ± 0.3
Wild type TSH receptor	6.1 ± 0.5 (0.04)	13.8 ± 1.0 (0.44)	25.0 ± 0.2
Asp 619 → Gly	5.3 ± 0.5 (0.00)	13.7 ± 0.9 (0.39)	19.8 ± 0.1
Ala 623 → Ile	5.9 ± 0.1 (0.03)	21.3 ± 0.6 (0.97)	25.2 ± 0.5

COS-7 cells were cotransfected with the various TSH receptor constructs and plasmid pXGH5. They were prepared and handled as described in Table 1, except that ³H-inositol (myo-[2-³H]inositol, 30 μ Ci ml⁻¹, 16.5 Ci mmol⁻¹; Dupont-NEN, Haren, Belgium) was added to the culture medium for the last 24 h of culture. After a 30-min preincubation in KRH buffer containing 10 mM LiCl, cells were further incubated for 15 min in the same medium containing TSH or ATP as indicated. Incorporation of ³H-inositol in inositol phosphates was determined according to ref. 24. It is expressed as 10³ × the ratio of the sum of ³H c.p.m. eluted in the inositol mono-, di- and trisphosphate fractions to the total c.p.m. incorporated (means of triplicate transfections $\pm$ s.e.m.). Values in parenthesis were obtained by subtracting the incorporation in cells transfected with pSVL vector alone, and dividing by the human growth hormone production to account for transfection efficiency (see Table 1 legend).

variant sequences and confirms the qualitative difference between quiescent and adenomatous tissue and the clonal nature of the lesions^{14,15}.

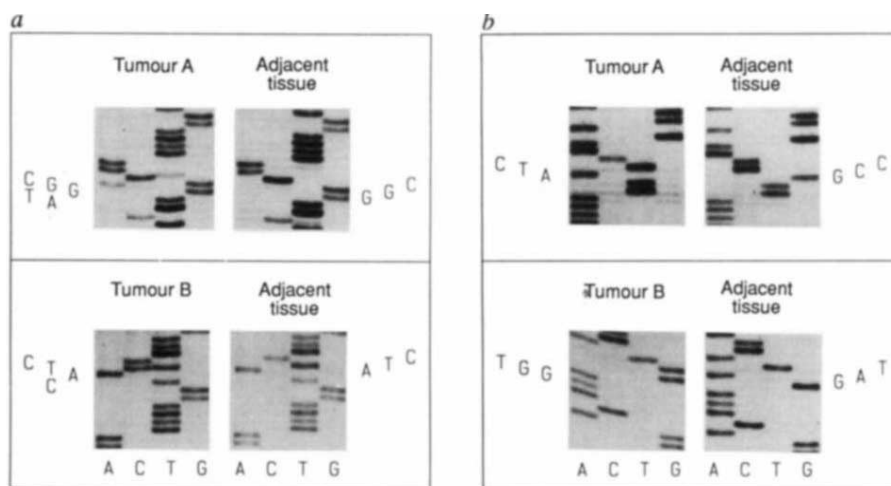
When primary structures of receptors are aligned to maximize sequence similarity, the Ala 623 → Ile mutation affects a residue of the TSH receptor that is the homologue of Ala 293, the mutation of which leads to constitutive activation of phospholipase C in the α_{1b} adrenergic receptor¹¹. The same mutation (Ala 623 → Ile) in the TSH receptor might constitutively activate adenyl cyclase, but mutants containing Glu or Lys in place of Ala 623 lose their ability to couple to G_q, with no effect on G_s-dependent activation of adenyl cyclase¹⁶. Although all mutations of Ala 293 have qualitatively similar effects in the α_{1b} adrenergic receptor¹¹, substitution of alanine by glutamic acid or lysine in the TSH receptor might have a different effect from the more conservative Ala → Ile replacement (see later).

The occurrence of the same Asp 619 → Gly mutation in two independent tumours indicates that it may be the cause of the disease. Asp 619 corresponds to an invariant residue in the highly conserved third cytoplasmic loop of glycoprotein hormone receptors. A glutamic acid residue is found at the corresponding position in virtually all receptors to small non-peptidic ligands and in the opsins¹⁷. A Glu 268 → Gly mutation in the β_2 adrenergic receptor, which is the homologue of the Asp 619 → Gly mutation in the TSH receptor, gave an increased affinity for isoprenaline, with a proportionate increase in sensitivity to the agonist¹⁸.

The actual effects of the mutations on receptor function were investigated by transfection in COS-7 cells. Cells transfected with Ala 623 → Ile or Asp 619 → Gly constructs showed higher basal accumulation of cAMP than those transfected with the wild-type receptor (Table 1). Their response to TSH was maintained, although it was lower when expressed relative to basal cAMP accumulation. Contrary to results with the mutant α_{1b} adrenergic receptor¹¹, the potency of TSH does not seem to be higher in the mutants. Investigation of the effects of both mutations on the activation of the phospholipase C cascade showed that there was no change in basal accumulation of inositol phosphates and

FIG. 1 Nucleotide sequence of the gene segment encoding the third intracellular loop of the TSH receptor (residues 589 to 647) in three hyperfunctioning adenomas and in adjacent tissue. *a*, Direct sequencing by PCR of non-coding strands, demonstrating the coexistence in the tumours of wild-type and mutated alleles. *b*, Sequences of the coding strands of the mutated alleles after subcloning in pSVL. In tumour A, codon 623 (GCC; Ala) is mutated to (ATC; Ile). In tumour B (and tumour C, data not shown) codon 619 (GAT; Asp) is mutated to (GGT; Gly).

METHODS. DNA was extracted²⁵ from tumours and extratumoural tissues obtained at surgery. The identity of samples originating from the same patient was checked by DNA fingerprinting²⁶. A fragment (from nucleotides 1,762 to 1,976) was amplified by PCR using a couple of primers of which the forward primer was biotinylated: forward primer 5'-TGTAACGACGCGCCAGTTATTGTT-TTTGTTCTGACGC-3', and reverse primer 5'-AACAGCTATGACCATGTGAG-AGGCTGTTCAGAATT-3'; underlined sequences correspond to M13 sequencing primers. PCR was done in a final volume of 100 µl containing 1 µg DNA, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 2 mM MgCl₂, 0.01% gelatine, 0.2 mM dNTPs, 5 U *Taq* polymerase (Cetus) and 150 nM of each primer; after denaturation at 93 °C for 2 min 30 s, 30 cycles (93 °C



for 1 min, 52 °C for 1 min, 72 °C for 30 s) were followed by a final elongation step of 6 min at 72 °C. PCR products were purified on strept-avidin-coated magnetic beads (Dynal, Oslo) and sequenced with Sequenase (version 2.0; USB Cleveland) and ³⁵S-labelled dATP-αS (Amersham, UK). PCR-amplified segments were subcloned in pSVL and the individual alleles were similarly sequenced.

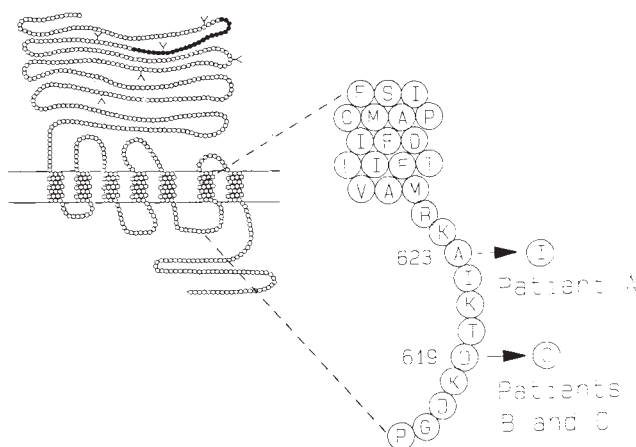


FIG. 2 Schematic representation of the TSH receptor. The third intracytoplasmic loop is enlarged to show the two activating mutations identified in our study.

that the inositol phosphate response to stimulation by TSH was maintained (Table 2). This is supported by the phospholipase C cascade being activated only when the TSH receptor is hyperstimulated².

Our results show that G-protein-coupled receptors are effectively capable of acting as proto-oncogenes¹¹. In tissues in which cAMP stimulates function and growth, these receptors stimulate adenylyl cyclase constitutively and will, as predicted^{3,19}, induce hyperfunctioning adenomas. In other cases, the gain-of-function caused by the activating mutations may be responsible for more subtle phenotypes, as described for the mouse melanocyte-stimulating hormone receptor²⁰. □

Received 30 April; accepted 9 September 1993.

- Dumont, J. E., Vassart, G. & Refetoff, S. In *The Metabolic Basis of Inherited Diseases* (ed. Scriver C.R.) 1843-1879 (McGraw-Hill, USA, 1989).
- Dumont, J. E., Lamy, F., Roger, P. & Maenhaut, C. *Physiol. Rev.* **72**, 667-697 (1992).
- Dumont, J. E., Jauniaux, J. C. & Roger, P. *Trends biochem. Sci.* **14**, 67-71 (1989).
- Weintraub, B. D., Gershengorn, M. C., Kourides, I. A. & Fein, H. *Ann. intern. Med.* **95**, 339-351 (1981).
- Zakarija, M. & McKenzie, J. M. *Exp. clin. Endocr.* **97**, 165-169 (1991).
- Lyons, J. et al. *Science* **249**, 655-659 (1990).
- Weinstein, I. S. et al. *New Engl. J. Med.* **325**, 1688-1695 (1991).
- O'Sullivan, C., Barton, C. M., Staddon, S. L., Brown, C. L. & Lemoine, N. R. *Molec. Carcinog.* **4**, 345-349 (1991).

- Suarez, H. G. et al. *Oncogene* **6**, 677-679 (1991).
- Kjelsberg, M. A., Cotechia, S., Ostrowski, J., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **267**, 1430-1433 (1992).
- Allen, L. F., Lefkowitz, R. J., Caron, M. G. & Cotechia, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11354-11358 (1991).
- Landis, C. A. et al. *Nature* **340**, 692-696 (1989).
- Van Sande, J. et al. *J. clin. Endocr. Metab.* **50**, 776-785 (1980).
- Van Sande, J. et al. *J. clin. Endocr. Metab.* **66**, 570-579 (1988).
- Thomas, G. A., Williams, D. & Williams, E. D. *Am. J. Path.* **134**, 141-151 (1989).
- Kosugi, S. *J. biol. Chem.* **267**, 24153-24156 (1992).
- Strosberg, A. D. *Eur. J. Biochem.* **196**, 1-10 (1991).
- O'Dowd, B. F. et al. *J. biol. Chem.* **263**, 15985-15992 (1988).
- Ruvkun, G. *Nature* **360**, 711-712 (1992).
- Robbins, L. S. et al. *Cell* **72**, 827-834 (1993).
- Libert, F. et al. *Biochem. biophys. Res. Commun.* **165**, 1250-1255 (1989).
- Christophe, D. et al. *Molec. cell Endocr.* **64**, 5-18 (1989).
- Brooker, G., Harper, J. F., Terasaki, W. L. & Moylan, R. D. *Adv. Cyclic Nucleotide Res.* **10**, 1-33 (1979).
- Berridge, M. J. *Biochem. J.* **212**, 849-858 (1983).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Vassart, G. et al. *Science* **235**, 683-684 (1987).

ACKNOWLEDGEMENTS. J.P. and L.D. contributed equally to this study. We thank C. Christophe for synthesizing oligonucleotides, C. Massaer for cAMP and inositol phosphate determination, and A. Flahaut for hGH determination. This study was supported by the Belgian Programme on University Poles of Attraction initiated by the Belgian State, Prime Minister's office, Science programming, and by grants from the Fonds de la Recherche Scientifique Médicale and the EEC programme for radioprotection. J.P. and L.D. are supported by Télévie and the Fondation Hoguet, respectively.

A constitutively activating mutation of the luteinizing hormone receptor in familial male precocious puberty

Andrew Shenker^{*†}, Louisa Laue[‡], Shinji Kosugi[§],
John J. Merendino Jr^{*}, Takashi Minegishi[¶]
& Gordon B. Cutler Jr[†]

^{*} Molecular Pathophysiology Branch, Building 10, Room 8C-101, NIDDK, [‡] Developmental Endocrinology Branch, NICHD, and [§] Cell Regulation Section, Laboratory of Biochemistry and Metabolism, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, USA

[¶] Department of Laboratory Medicine, Kyoto University School of Medicine, Kyoto 606, Japan

[†] Department of Obstetrics and Gynecology, Gunma University School of Medicine, Maebashi, Gunma 371, Japan

FAMILIAL male precocious puberty (FMPP) is a gonadotropin-independent disorder that is inherited in an autosomal dominant, male-limited pattern¹⁻⁵. Affected males generally exhibit signs of puberty by age 4. Testosterone production and Leydig cell hyperplasia occur in the context of prepubertal levels of luteinizing hormone (LH)³⁻⁵. The LH receptor is a member of the family of G-protein-coupled receptors^{6,7}, and we hypothesized that FMPP might be due to a mutant receptor that is activated in the presence of little or no agonist⁸⁻¹². A single A → G base change that results in substitution of glycine for aspartate at position 578 in the sixth transmembrane helix of the LH receptor was found in affected individuals from eight different families. Linkage of the mutation to FMPP was supported by restriction-digest analysis. COS-7 cells expressing the mutant LH receptor exhibited markedly increased cyclic AMP production in the absence of agonist, suggesting that autonomous Leydig cell activity in FMPP is caused by a constitutively activated LH receptor.

Genomic DNA was isolated from affected males from eight different FMPP families, and polymerase chain reaction (PCR) was used to amplify a fragment of LH receptor (LHR) DNA encoding amino-acid residues 441 to 594 (which includes most of transmembrane helices 3 to 6, the second extracellular loop, and the second and third intracellular loops)^{6,7}. PCR products were screened for heterozygous mutations using temperature-gradient gel electrophoresis (TGGE)¹³. PCR product from normal individuals migrated as a single band in a gradient of increasing temperature, but PCR products from all patients with FMPP migrated as multiple bands (Fig. 1a). The abnormal pattern of bands was identical for all patients.

PCR product from patient V-4 was subcloned and sequenced, revealing a heterozygous A → G transition at nucleotide 1,733 in codon 578 (Fig. 1b). This result was confirmed by direct sequencing of PCR products from patients from two other families. The mutation (GAT to GGT) encodes a substitution of glycine for an aspartate in transmembrane helix 6 (Fig. 1c) and creates a recognition site for the restriction endonuclease *MspI*. Restriction digests were positive for the Asp 578 → Gly mutation in all patients who had abnormal TGGE patterns, and analysis of PCR products from three generations of one well characterized FMPP kindred indicates linkage of the LHR mutation to FMPP (Fig. 2).

Although the eight FMPP families are not known to be related, six of them originate in the same geographical region, and the surname of the kindred shown in Fig. 2 appears in the line of descent of two other families. We suspect that there is a common ancestral origin for the Asp 578 → Gly mutation, but

that different mutations may be found in other families with FMPP.

To assess the functional effect of the Asp 578 → Gly mutation directly, wild-type and mutated human LHR were transiently expressed in COS-7 cells and cAMP accumulation was measured (Fig. 3). Cells transfected with wild-type LHR DNA had the same basal cAMP production as cells transfected with vector DNA. Human chorionic gonadotropin (hCG) produced a concentration-dependent increase in cAMP production in cells expressing the wild-type LHR, with an EC₅₀ (50% effective concentration) of 4 ng ml⁻¹ and mean maximal stimulation of 11.5-fold (range 2.9- to 18.7-fold; *n* = 5).

In contrast to the wild-type receptor, the mutant LHR produced a 4.5-fold increase in basal cAMP production in COS-7 cells (range 1.8- to 8.3-fold; *n* = 5), indicating that it was constitutively active. There was a significant increase in basal cAMP production even when cells were transfected with 25-fold less mutant DNA (data not shown). The mutant receptor was capable of responding to increasing concentrations of hCG, with an EC₅₀ similar to that of the wild-type receptor. Maximal hCG-stimulated cAMP production in mutant-transfected cells was variable, but did not differ significantly from that in wild-type-transfected cells (range 75 to 135% of wild-type maximum). Agonist-independent stimulation of cAMP production by mutant receptors represented 42% (range 32 to 62%; *n* = 5) of the maximal stimulation produced by hCG.

A dominant mutation that leads to constitutive activation of the LHR-mediated cAMP signalling pathway can explain the pathophysiology of FMPP, including its male-limited inheritance pattern. LHR-mediated effects, including testosterone production, involve increased production of cellular cAMP¹⁴. The normal onset of puberty in boys is attributed to a rise in circulating LH levels. Concentrations of LH found in the serum of pubertal boys¹⁵ produce less than half-maximal activation of normal, recombinant human LHR⁷. We postulate that intracellular cAMP accumulation triggered by the unoccupied mutant receptor is sufficient to cause Leydig cell hyperfunction and hyperplasia. The age of pubertal onset in FMPP may depend on the extent to which the mutant allele is expressed as protein and on the relative expression of other genes critical for Leydig cell maturation. Because LH is sufficient to trigger steroidogenesis in Leydig cells but both LH and FSH are required to activate ovarian steroidogenesis, inappropriate activation of LHR alone would not be expected to cause precocious puberty in females. For example, hCG-secreting germ-cell tumours cause sexual precocity in males, but not females¹⁶. The observation that plasma testosterone in boys with FMPP increases normally in response to high doses of exogenous hCG^{4,5} is consistent with the ability of the mutant LHR in COS cells to respond to high concentrations of hCG with the same maximal response as wild-type LHR (Fig. 3).

Other constitutively active G-protein-coupled receptors have been produced by mutating residues in the C-terminal portion of the third intracellular loop, adjacent to helix 6 (refs 8-12). A model of the arrangement of the transmembrane α -helices in G-protein-coupled receptors¹⁷ places Asp 578 in the middle of helix 6, oriented towards the internal cleft. The aspartic acid at position 578 is conserved in all glycoprotein hormone receptors, but is not found in any other G-protein-coupled receptor¹⁸. The corresponding residue in cone opsins (Phe or Tyr at residue 277) lies near the retinal chromophore and is critical in determining the spectral absorption properties that are the basis of colour vision¹⁹. A phenylalanine residue occupies the equivalent position in all cationic neurotransmitter receptors, and is postulated to participate in a network of interhelical, aromatic interactions which help form the agonist-binding pocket and may be involved in propagating a conformational signal to the intracellular loops¹⁸.

We speculate that in the inactive receptor state, Asp 578 is engaged in electrostatic or hydrogen-bond interaction with one

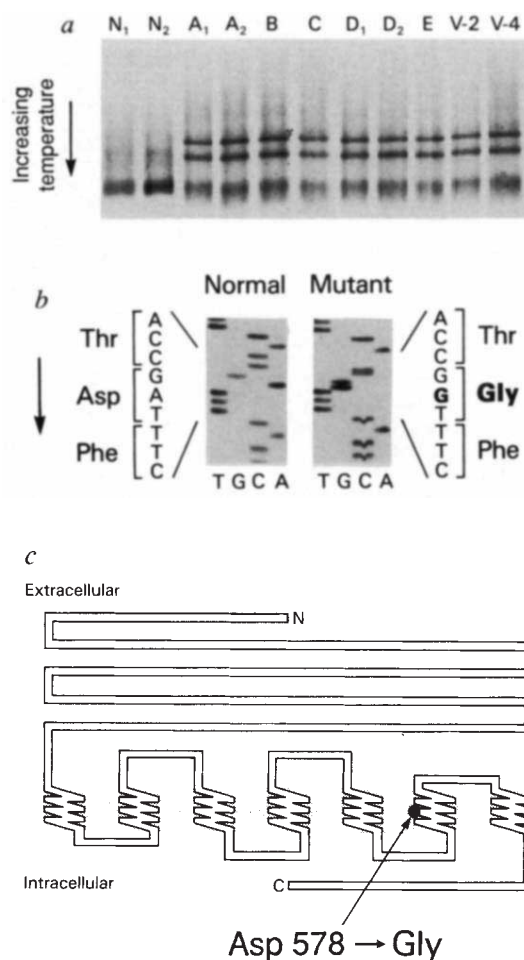
[†] To whom correspondence should be addressed.

or more residues in an adjacent helix, thus constraining the conformation of helix 6. When agonist binds to LHR these bonds are disrupted and a conformational change ensues. Substitution of the polar aspartate residue with glycine, a residue incapable of forming electrostatic or hydrogen bonds, would allow LHR

to assume a partially active conformation in the absence of agonist. The substitution appears to have a negligible effect on agonist affinity, which is determined primarily by binding to the large N-terminal extracellular domain²⁰, or on the ability of the receptor to be maximally activated by agonist.

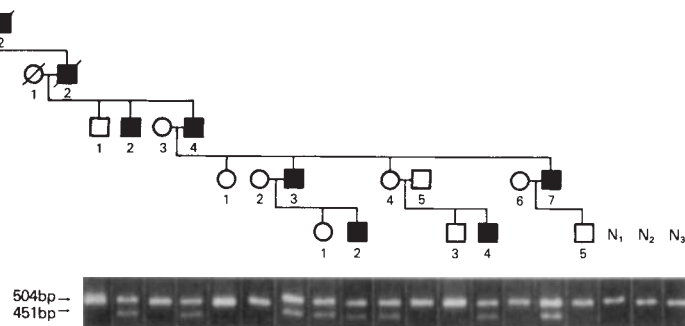
FIG. 1 a, Temperature-gradient gel electrophoresis (TGGE) of PCR-amplified DNA fragments encoding amino-acid residues 441 to 594 of the human LHR from 2 normal males (N_1 and N_2) and from 9 FMPP patients from 6 different families. TGGE identifies DNA duplexes containing base-pair mismatches on the basis of their altered melting properties in comparison to wild-type DNA duplexes^{13,28}. The abnormal pattern of bands produced by each patient sample consists of the wild-type homoduplex band, two DNA heteroduplexes that melt at a lower temperature than the wild-type band, and a poorly focused mutant homoduplex band that melts at a slightly higher temperature than the wild-type band. This suggested that the DNA fragments from FMPP patients contained a heterozygous substitution of a G-C for an A-T base pair. DNA samples from patients from 2 other families produced the same abnormal pattern of bands (data not shown). **b**, Subcloned PCR product from subject V-4, showing normal (left) and mutant (right) sequences. Four clones had wild-type LHR sequence, and three clones showed an A-to-G transition at nucleotide 1,733. The heterozygous mutation encodes substitution of Asp 578 with Gly. The base change causes a compression of the 2 adjacent cytosines in the sequencing ladder. **c**, Schematic representation of the membrane topology of the human LHR, indicating the position of Asp 578 in the middle of the sixth transmembrane helix.

METHODS. The study was approved by the Clinical Research Committee of the NICHD and informed consent was obtained for all participants. Subject A_1 was patient 8 in ref. 29 and is the third cousin of subject A_2 ; subject D_1 is the son of subject D_2 ; subject E was patient 2 in other studies^{29,30} and is the son of patient C.C. in ref. 2; subjects V-2 and V-4 are cousins (see Fig. 2 pedigree); subject V-4 was patient 1 in ref. 30 and patient 5 in ref. 29. PCR primers were selected based on the predicted melting behaviour of GC-clamped LHR DNA fragments (EZMELT, The Telluride Group, Newton, Massachusetts; and MELT87, L. Lerman, MIT). A GC clamp facilitates TGGE detection of single base substitutions within a DNA fragment²⁸. Primers 5'-CGCCCCGCGCGCCCCGCGCCCCGCGCCCCGCGCCCCGCGCCCCGCGCTGCTG-GCTTTTTCACGTGATT3' and 5'-TGAAGGAGCTGAGATGGCAAAA3' generate a 504-bp DNA fragment consisting of a 40-bp GC clamp (underlined) and nucleotides 1,320 to 1,783 of the LHR sequence^{6,7}. Each PCR reaction mixture (100 μ l) contained 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.001% gelatin, 25 pmol each primer, 2.0 mM MgCl₂, 200 μ M each nucleoside triphosphate, 2.5 U Taq polymerase (Perkin Elmer-Cetus), and 5 μ l (~1 μ g) of crude genomic DNA prepared from fresh or frozen blood samples. PCR conditions were: 5 min at 95 °C, 30 cycles of 1 min at 57 °C, 30 s at 72 °C and 1 min at 95 °C, and a final extension of 3 min at 72 °C. A commercial apparatus (Diagen) was used for TGGE. Gels were 3% polyacrylamide (50:1 acrylamide:bisacrylamide), 8 M urea, 1 mM EDTA, 2% glycerol, 20 mM MOPS-NaOH, pH 8.0. PCR fragments were denatured (at 95 °C for 5 min) and reannealed (50 °C, 15 min) in loading buffer (20 mM MOPS-NaOH, pH 8.0, 5 mM EDTA,



0.1% bromophenol blue, 2% glycerol). Each lane had 1 μ l of PCR reaction loaded and samples were electrophoresed in 20 mM MOPS-NaOH, 1 mM EDTA, pH 8.0, at 280 V for 75 min over a linear temperature gradient of 35 to 55 °C. DNA was visualized using silver staining. PCR product generated from subject V-4 genomic DNA was blunt-end-ligated into the *Sma*I site of pBluescript (Stratagene); plasmid subclones were sequenced by standard methods.

FIG. 2 *Msp*I restriction enzyme analysis of DNA from members of a family with FMPP and 3 unrelated normal males (N_1 , N_2 , N_3). Affected males are shown as solid squares. Subjects I-2 and II-2 were classified as affected on the basis of short stature. An A-to-G mutation at nucleotide 1,733 of the LHR gene creates the recognition site for *Msp*I (C/CGG). Normal PCR product (504 bp) will not be cut by *Msp*I, but DNA containing the mutant sequence is digested into two fragments of 451 and 53 bp. DNA from subjects who are heterozygous for the mutation will yield both uncut and digested fragments. The 53-bp fragment was run off the bottom of the gel in this experiment. The proband (V-4), his cousin (V-2), all other affected males in the pedigree (III-2, III-4, IV-3, IV-7), and an obligate carrier female (IV-4) showed the heterozygous digest pattern, whereas unaffected males (III-1, IV-5, V-3) and unrelated normal males (N_1 , N_2 , N_3) did not, suggesting linkage of the LHR mutation to FMPP. Restriction analysis indicates that the DNA of female IV-1 is normal, but that her niece (V-1) carries the Asp 578 → Gly mutation. Amniocyte DNA from a male fetus (V-5) at risk for inheriting FMPP was



normal. DNA from affected individuals in the 7 other FMPP families showed the heterozygous digest pattern, but 40 additional control DNA samples did not (data not shown).

METHODS. PCR products (Fig. 1 legend) were digested for 16 h with 40 units of *Msp*I (New England Biolabs), separated on a 6% polyacrylamide gel (Novex), and visualized with ethidium bromide.

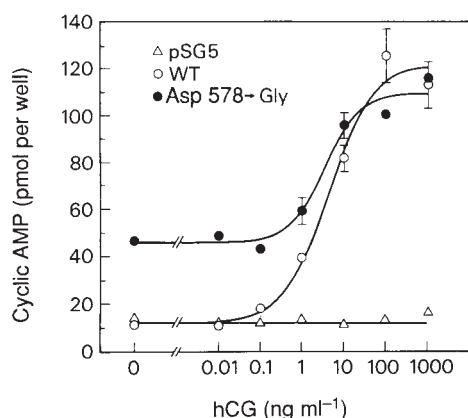


FIG. 3 Basal and human chorionic gonadotropin (hCG)-stimulated cAMP accumulation in COS-7 cells transfected 48 h earlier with pSG5 vector alone (pSG5) wild-type LHR DNA (WT), or mutant LHR DNA (Asp 578 → Gly). Data points are mean $\pm$ s.e. from 3 (hCG) or 6 (basal) replicate wells from a single experiment. If not shown, s.e. is smaller than the symbol. Results were similar in 4 additional experiments.

METHODS. Methods for mutagenesis, transfection and cAMP assay have been described¹². Human LHR cDNA was inserted into a M13mp18 vector and oligonucleotide-mediated site-directed mutagenesis was used to generate clones encoding the Asp 578 → Gly mutation (T7GEN kit, US Biochemical). *Hae*III was used instead of *Msp*I for parental strand digestion. Wild-type and mutant clones were inserted into the *Eco*RI site of the pSG5 expression vector (Stratagene). Mutagenesis was confirmed by DNA sequencing of the final construct and plasmid DNA was purified by CsCl gradient ultracentrifugation. COS-7 cells ($\sim 10^7$ cells) were transfected by electroporation (Bio-Rad) with 25 μ g plasmid DNA, suspended in Dulbecco's modified Eagle's medium containing 10% fetal calf serum, and transferred to 24-well plates ($\sim 10^5$ cells per well). Equivalent transfection efficiency was confirmed by co-transfecting pSVGH and measuring growth hormone concentration in the medium. Forty-eight hours after transfection, cells were washed and then incubated for 1 h at 37 °C with 0.2 ml Hanks' balanced salt solution containing 0.5% bovine serum albumin, 20 mM HEPES, pH 7.4, 10 mM LiCl and 0.5 mM 3-isobutyl-1-methylxanthine alone or with 0.01 to 1,000 ng ml⁻¹ hCG (CR-127, 14,900 IU mg⁻¹; National Hormone and Pituitary Program). Perchloric acid was added to each well, samples were centrifuged, and total cAMP in aliquots of neutralized supernatant was determined by ¹²⁵I radioimmunoassay (DuPont). Membrane protein and growth hormone concentration per well varied by less than 10% between wild-type and mutant transfections.

Constitutively activating point mutations have been described in other classes of transmembrane receptors^{21,22}. Inactivating mutations of G-protein-coupled receptors can serve as a mechanism of human disease^{23,24}, and the identification of amino-acid substitutions that cause constitutive activation of adrenergic receptors *in vitro* led to the prediction that such mutations might also be pathogenic^{8,9}. Activating mutations of the melanocyte-stimulating hormone receptor have been found in mice with dominantly inherited hyperpigmentation²⁵, and activating mutations of rhodopsin have been described in a severe form of retinitis pigmentosa²⁶ and in congenital stationary night-blindness²⁷. FMPP provides the first example of an inherited human disease that is due to a constitutively activating mutation in a G-protein-coupled hormone receptor. □

Received 30 June; accepted 9 September 1993.

- Stone, R. K. *Am. J. med. Sci.* **24**, 561–564 (1852).
- Walker, S. H. *J. Pediatr.* **41**, 251–257 (1952).
- Schedewie, H. K. et al. *J. clin. Endocr. Metab.* **57**, 271–278 (1981).
- Rosenthal, S. M., Grumbach, M. M. & Kaplan, S. L. *J. clin. Endocr. Metab.* **57**, 571–578 (1983).
- Holland, F. J. *Endocr. Metab. Clin. North Am.* **20**, 191–210 (1991).
- Minegishi, T. et al. *Biochem. biophys. Res. Commun.* **172**, 1049–1054 (1990).
- Jia, X.-C. et al. *Molec. Endocr.* **5**, 759–768 (1991).
- Allen, L. F., Lefkowitz, R. J., Caron, M. G. & Cotechia, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11354–11358 (1991).
- Kjelsberg, M. A., Cotechia, S., Ostrowski, J., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **267**, 1430–1433 (1992).

- Samama, P., Cotechia, S., Costa, T. & Lefkowitz, R. J. *J. biol. Chem.* **268**, 4625–4636 (1993).
- Ren, Q., Kurose, H., Lefkowitz, R. J. & Cotechia, S. *J. biol. Chem.* **268**, 16483–16487 (1993).
- Kosugi, S., Okajima, F., Ban, T., Hidaka, A., Shenker, A. & Kohn, L. D. *Molec. Endocr.* **7**, 1009–1020 (1993).
- Riesner, D. et al. *Electrophoresis* **10**, 377–389 (1989).
- Hunzicker-Dunn, M. & Birnbaumer, L. in *Luteinizing Hormone Action and Receptors* (ed. Ascoli, M.) 57–134 (CRC, Boca Raton, 1985).
- Oerter, K. E., Uriarte, M., Rose, S. R., Barnes, K. M. & Cutler, G. B. Jr *J. clin. Endocr. Metab.* **71**, 1251–1258 (1990).
- Sklar, C. A., Conte, F. A., Kaplan, S. L. & Grumbach, M. M. *J. clin. Endocr. Metab.* **53**, 656–660 (1981).
- Baldwin, J. M. *EMBO J.* **12**, 1693–1703 (1993).
- Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels, A. & Hibert, M. J. *Med. Chem.* **35**, 3448–3462 (1992).
- Neitz, M., Neitz, J. & Jacobs, G. J. *Science* **252**, 971–974 (1991).
- Tsai-Morris, C. H., Buczek, E., Wang, W. & Dufau, M. L. *J. biol. Chem.* **265**, 19385–19388 (1990).
- Bargmann, C. I., Hung, M.-C. & Weinberg, R. A. *Cell* **45**, 649–657 (1986).
- Yaghamai, R. & Hazelbauer, G. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7890–7894 (1992).
- Dryja, T. P. et al. *New Engl. J. Med.* **323**, 1302–1307 (1990).
- Rosenthal, W. et al. *Nature* **359**, 233–235 (1992).
- Robbins, L. S. et al. *Cell* **72**, 827–834 (1993).
- Robinson, P. R., Cohen, G. B., Zhukovsky, E. A. & Oprian, D. D. *Neuron* **9**, 719–725 (1992).
- Dryja, T. P., Berson, E. L., Rao, V. R. & Oprian, D. D. *Nature Genet.* **4**, 280–283 (1993).
- Myers, R. M., Sheffield, V. C. & Cox, D. R. in *PCR Technology: Principles and Applications for DNA Amplification* (ed. Erlich, H. A.) 71–88 (Stockton, New York, 1989).
- Laue, L., Jones, J., Barnes, K. M. & Cutler, G. B. Jr *J. clin. Endocr. Metab.* **76**, 151–155 (1993).
- Laue, L. O. H. et al. *New Engl. J. Med.* **320**, 496–502 (1989).

ACKNOWLEDGEMENTS. We thank A. M. Spiegel, L. D. Kohn and T. Mori for their support; A. Garritsen, P. J. M. van Galen, A. Hidaka, D. Accili, S. Wank and M. Dufau for discussion; and also the FMPP families, J. Jones, L. Lerman, and the National Hormone and Pituitary Program.

Prevention of lung reperfusion injury in rabbits by a monoclonal antibody against interleukin-8

Nobuaki Sekido[†], Naofumi Mukaida^{*}, Akihisa Harada[†], Isao Nakanishi[§], Yoh Watanabe[†] & Kouji Matsushima^{*} ||

^{*} Department of Pharmacology, Cancer Research Institute, [†] First Department of Surgery, [‡] Department of Hygiene and [§] First Department of Pathology, School of Medicine, Kanazawa University, 13-1 Takara-machi, Kanazawa, Ishikawa 920, Japan

RE-ESTABLISHING blood flow to ischaemic tissues causes greater injury than that induced during the ischaemic period^{1,2}. This type of tissue injury, reperfusion injury, is involved in frostbite, multiple organ failure after hypovolaemia and in myocardial infarction¹. Depletion of neutrophils alleviates reperfusion injury, implying a causal role of neutrophil infiltration^{3,4}. Among members of the recently discovered family of chemotactic cytokines (chemokines)^{5–8}, interleukin-8 (IL-8)^{5,9–13} is a major neutrophil chemotactic and activating factor produced by various types of human cells. We investigated its pathophysiological role in a rabbit model of a lung reperfusion injury. Reperfusion of ischaemic lung caused neutrophil infiltration and destruction of pulmonary structure, as well as local production of IL-8. Furthermore, the administration of a neutralizing monoclonal antibody against IL-8 prevented neutrophil infiltration and tissue injury, proving a causal role of locally produced IL-8 in this model.

In acute inflammation, neutrophils infiltrate down a concentration gradient of chemotactic factor(s) after adherence to endothelial cells through adhesion molecules¹⁴. Antibodies against those adhesion molecules inhibited the reperfusion injury^{15–17}, suggesting their essential role in the injury. Leukotriene B₄ (LTB₄) was produced in reperfusion injuries^{18–21} but the effects of its specific antagonists on the injury were

|| To whom all correspondence should be addressed.

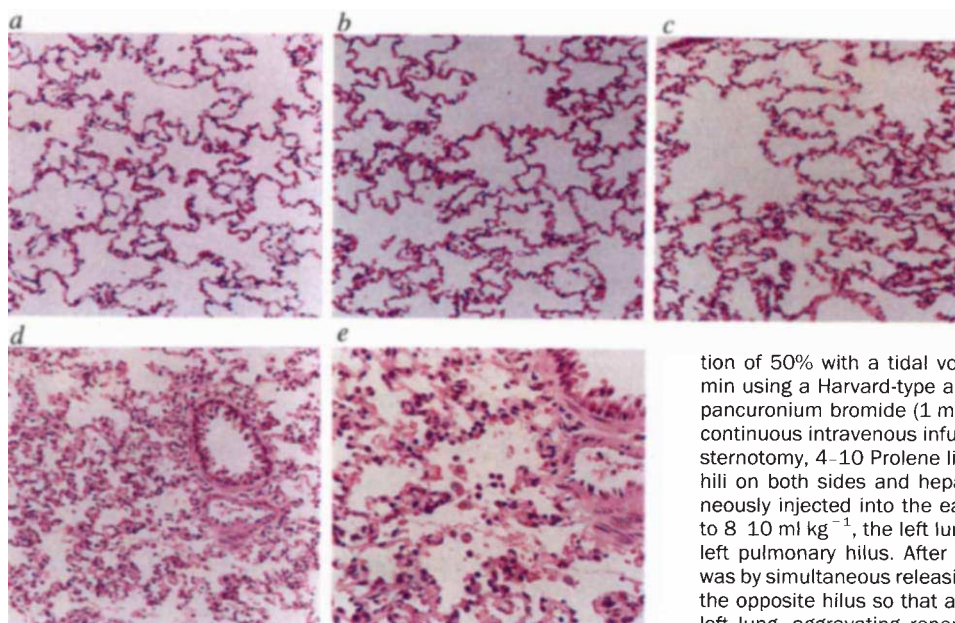


FIG. 1 Histological examination of lung tissues; a, untreated; b, after 5 h ischaemia; c, 2 h ischaemia + 2 h reperfusion; d, e, 2 h ischaemia + 3 h reperfusion. Haematoxylin–eosin stained sections at $\times 46$ (a–d) or $\times 92$ (e) magnification.

METHODS. New Zealand White female rabbits (body weight 2.5–3.0 kg) were anaesthetized by intravenous administration of 25 mg kg^{-1} sodium pentobarbital. Rabbits were then ventilated at an O_2 concentra-

tion of 50% with a tidal volume of 15–20 ml kg^{-1} at 30 breaths per min using a Harvard-type animal respirator. Animals were relaxed with pancuronium bromide (1 mg) intravenously and kept anaesthetized on continuous intravenous infusion of ketamine (10 mg h^{-1}). After median sternotomy, 4–10 Prolene ligatures were passed around the pulmonary hilum on both sides and heparin sodium (200 units kg^{-1}) was simultaneously injected into the ear vein. After the tidal volume was reduced to 8–10 ml kg^{-1} , the left lung was rendered ischaemic by clamping the left pulmonary hilum. After ischaemia for 2 h, reperfusion to left lung was by simultaneous releasing of clamping of the left hilum and clamping the opposite hilum so that all pulmonary blood flow was directed to the left lung, aggravating reperfusion injury. Hyperoxygenation of the left lung was by clamping the opposite hilum. At the indicated time intervals, animals were killed by intravenous administration of an overdose of pentobarbital after collecting plasmas and BALF. Lung tissues were fixed in 10% formalin followed by haematoxylin–eosin staining. All the experiments were repeated at least three times using different rabbits and representative data are shown. All the procedures of animal experiments complied with the standards set out in the *Guideline for the Care and Use of Laboratory Animals in Takara-machi Campus of Kanazawa University*.

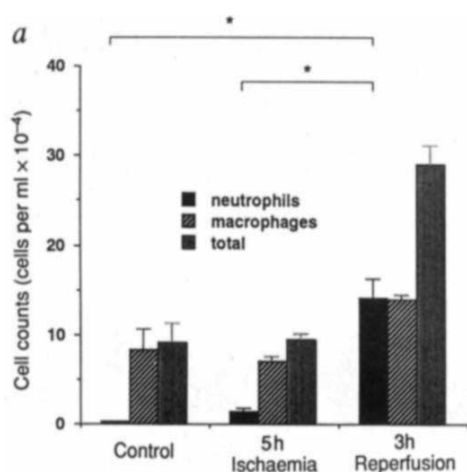
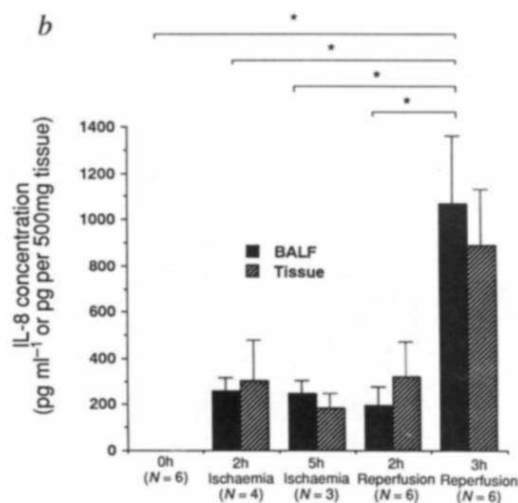


FIG. 2 a, The number of neutrophils (filled bars) and macrophages (hatched bars) as well as total cell number (dotted bars) was determined in BALF from either untreated, 5 h-ischaemic, or 2 h-ischaemic + 3 h-reperfused rabbit lungs. The mean and 1 s.d. for three rabbits are shown here. Statistical analysis was one-way analysis of variance (ANOVA) by Scheffe's multiple comparison procedure, and significance with $P < 0.05$ is indicated by asterisks. b, IL-8 contents in BALF (pg ml^{-1} ; filled bars) and in homogenates of lung tissues (pg per 500 mg tissue hatched bars) were shown. The number of animals for each group are indicated in parenthesis. Statistical significance was determined as described for a, and significance with $P < 0.05$ is indicated by asterisks.

METHODS. At the indicated time, BALF was recovered from the left main bronchus by injecting 5 ml physiological saline containing 1 mg EDTA. Recovered volume of BALF was 2.2 to 2.8 ml in all the experiments. BALF was then centrifuged at 1,500 r.p.m. for 15 min and the supernatants collected and stored at -80°C until enzyme-linked immunosorbent assay (ELISA). Cell pellets were resuspended in 100 ml



RPMI 1640 medium. Cell number was counted using a Burker-Turk haemocytometer and the proportions of cells were determined by differential counting of 200–400 cells on Giemsa-stained cell suspension smears. Lung tissues (500 mg) was homogenized with phosphate-buffered saline (PBS) containing 0.5% bovine serum albumin and 0.05% Tween 20 using a syringe. Homogenates were centrifuged at 3,000 r.p.m. for 15 min and the resultant supernatants collected and stored at -80°C until ELISA. We purified recombinant rabbit IL-8 expressed in *Escherichia coli* and immunized a guinea-pig with this recombinant rabbit IL-8. ELISA for rabbit IL-8 was established combining monoclonal anti-IL-8 (WS-4, 0.5 $\mu\text{g ml}^{-1}$), guinea-pig anti-rabbit IL-8 (1 $\mu\text{g ml}^{-1}$), and alkaline phosphatase-conjugated anti-guinea-pig IgG (diluted at 1/3,000; BioMakor, Rehovot, Israel) antibodies as capture, second and detection ones, respectively. Recombinant rabbit IL-8 was used as a standard and enzyme reaction was as previously described²⁸. The sensitivity of this ELISA is 50 pg ml^{-1} .

inconsistent^{19,21}. Hyperoxia after anoxia induced human peripheral blood mononuclear cells to produce IL-8 (ref. 22), suggesting IL-8 production on reperfusion. But low serum IL-8 levels were detected transiently in acute myocardial infarction²³. Intravenous injection of a large dose of IL-8 prevented tissue damage as well as neutrophil infiltration in reperfusion injury of rabbit heart²⁴. This might disrupt a concentration gradient of IL-8 by increasing blood IL-8 level above tissue level, or blocking

adhesion of neutrophils to endothelium²⁵. But this type of effect is not restricted to IL-8²⁶. Thus, we have tried to define the role of IL-8 in the lung reperfusion injury in rabbits by using a neutralizing monoclonal anti-IL-8 antibody.

Minimal morphological changes were detected in lungs rendered ischaemic for 5 hours (Fig. 1b) or hyperoxygenated for 5 hours (data not shown). In contrast, moderate infiltration of macrophages and neutrophils into the alveoli with focal fibrin-

FIG. 3 Immunohistochemical analyses on lung tissue reperfused for 3 h (a, b and d), and rendered ischaemic for 2 h (c). After immunostaining using either monoclonal anti-IL-8 antibody, WS-4, (a, b and c)²⁸ or the control monoclonal antibody with the same isotype, TpM-1 (d)²⁹, sections were observed at $\times 46$ (a, c and d) or $\times 92$ magnification (b).

METHODS. The paraffin-embedded sections of lung tissue were deparaffinized and dehydrated by sequential immersion into xylene, absolute ethanol, 70% ethanol and PBS. The sections were then immunostained by the labelled streptavidin-biotin method³¹. In brief, after incubation with anti-IL-8 (WS-4) or control antibody (TpM-1) at a concentration of $40 \mu\text{g ml}^{-1}$, sections were treated sequentially with biotinylated anti-mouse IgG and streptavidin-conjugated peroxidase (Nichirei, Tokyo, Japan). Immune complexes were detected using a peroxidase substrate (Kyowa Medics, Tokyo, Japan) consisting of diaminobenzidine-tetrahydrochloride for 2 min. Slides were then counterstained with methyl green.

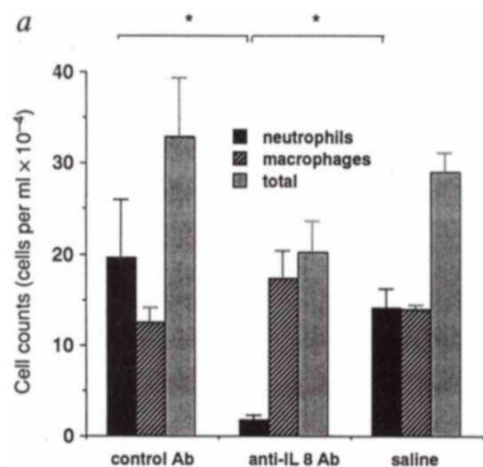
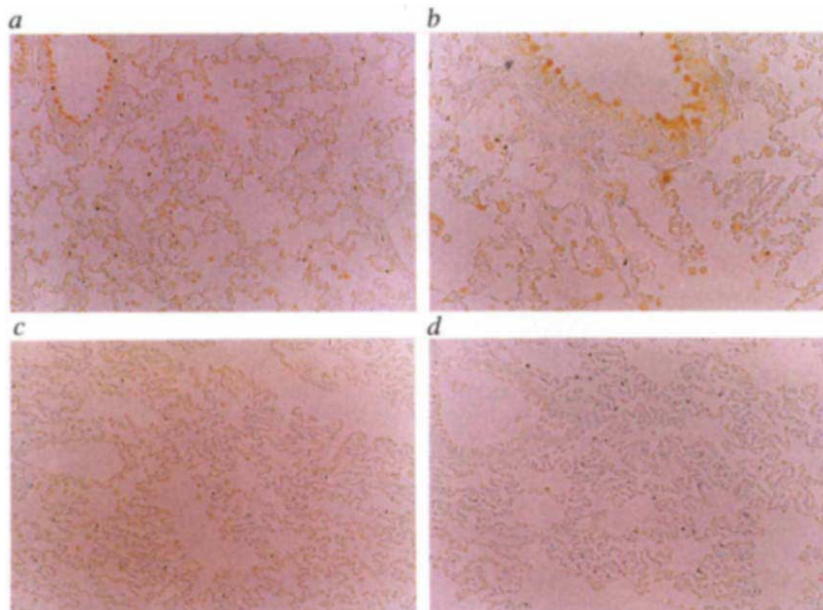
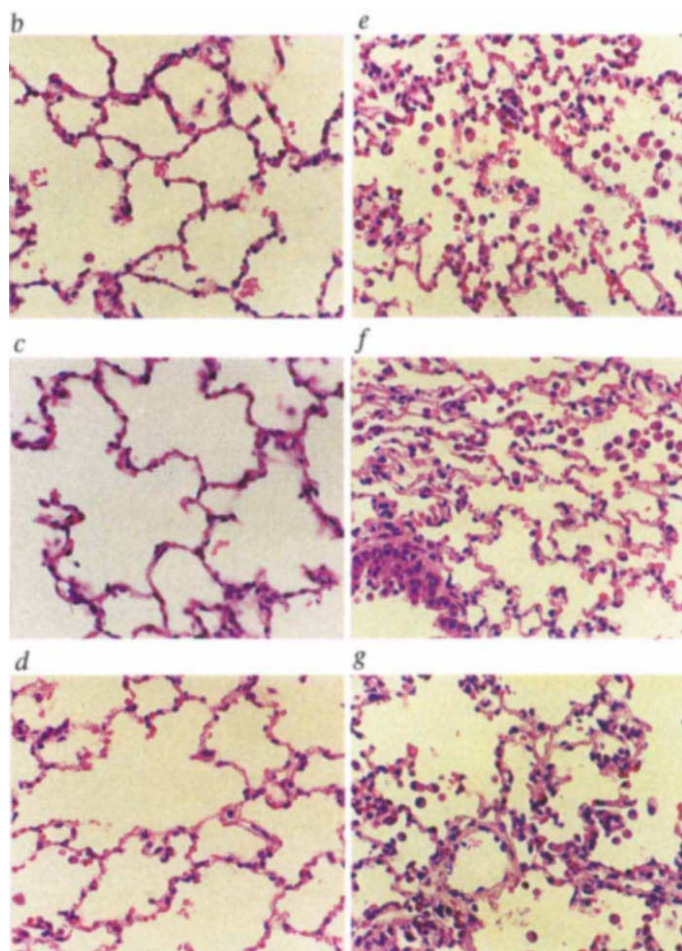


FIG. 4 a, Effects of anti-IL-8 on cell compositions of BALF. The number of neutrophils (filled bars) and macrophages (hatched bars) as well as total cell number (dotted bars) was determined in BALF from rabbits reperfused for 3 h after administration of either saline, anti-IL-8, WS-4, and a control antibody, TpM-1, respectively. Three rabbits were used for each group and the mean and standard deviation calculated on each group are shown here. Statistical analysis was as described in the legend to Fig. 2a, and significance with $P < 0.05$ is indicated by asterisks. b–g, Microscopic examination of effects of anti-IL-8 (b–d) and control antibody (e–g) on the reperfused lung tissues stained with haematoxylin–eosin stain ($\times 92$ magnification). Histological examinations were done on 3 rabbits for each group and representative results from individual animals are shown.

METHODS. Anti-IL-8 (WS-4) or control antibody (TpM-1) (10 mg) was dissolved in 3 ml saline. Antibodies were administered intravenously at the beginning of reperfusion. BALF and lung tissues were collected and processed for determination of cell components and histological examination, respectively, as described.



ous exudate was observed 2 h after reperfusion (Fig. 1c). Three hours after reperfusion, neutrophils markedly infiltrated into alveolar septa and alveoli with intravascular neutrophil aggregation (Fig. 1d, e). Extensive oedema and fibrinous exudates in alveolar lumina were observed as well as destruction of pulmonary structure. The number of neutrophils in bronchoalveolar lavage fluid (BALF) increased markedly at 3 h, whereas macrophages increased less (Fig. 2a), indicating the recruitment of primarily neutrophils into the injury site.

IL-8 was barely detectable in BALF or lung tissue homogenates before treatment (Fig. 2b). Ischaemia for 2 h increased IL-8 contents in both BALF and tissue homogenate, although not significantly. IL-8 contents did not increase further even when ischaemia continued for 5 h or if lungs were hyperoxygenated for 5 h (data not shown). Reperfusion to ischaemic lungs for 3 h, however, significantly increased IL-8 contents in both BALF and tissue homogenate without noticeable elevation of plasma IL-8 level. These findings suggest that a concentration gradient existed, attracting neutrophils from the bloodstream into injury sites, or alternatively that locally produced IL-8 bound to the surface of the endothelium, promoting migration of adherent neutrophils²⁷.

Local production of IL-8 was further verified by immunohistochemical analyses, as IL-8 protein was detected mainly in the cytoplasm and/or surface of bronchiolar ciliary cells and alveolar macrophages of lungs reperfused for 3 h (Fig. 3a, b). In contrast, lung tissues rendered ischaemic for 2 h (Fig. 3c) did not contain detectable amounts of IL-8. Lung tissue reperfused for 3 h was not stained with a control antibody (Fig. 3d), indicating the specificity of the reaction.

A monoclonal anti-human IL-8 antibody, WS-4 (ref. 28), inhibited the binding to rabbit neutrophils and neutrophil chemotactic activity of recombinant rabbit IL-8 at a WS-4: IL-8 molar ratio of nearly 1:1. Although a crossreactivity against other rabbit chemokines could not be completely excluded, WS-4 did not show any crossreactivity against any of several members of human chemokines including gro, IP-10, β -thromboglobulin, platelet factor 4, and monocyte chemotactic and activating factor. Intravenous administration of anti-IL-8 inhibited the increase of neutrophils in BALF, whereas the control antibody²⁹ did not (Fig. 4a). Anti-IL-8 also markedly decreased fibrinous exudation in the alveolar lumina, destruction of alveolar architecture, and neutrophil infiltration in all three rabbits examined histologically (Fig. 4b–d) whereas the control antibody failed (Fig. 4e–g), indicating a crucial role of IL-8 or other chemokine(s) which is immunochemically related to IL-8 in this model of lung reperfusion injury.

Although generation of reactive oxygen metabolites is a prominent biochemical change in reperfusion injury, little is known about how it causes its pathological hallmark, neutrophil infiltration^{1,2}. Because several oxygen radical scavengers inhibited IL-8 production³⁰ and a well-known free-radical generator, paraquat, induced IL-8 production (our unpublished data), reactive oxygen metabolites generated on reperfusion may induce IL-8 production. IL-8, in turn, promotes tissue damage by inducing neutrophil infiltration as well as release of lysosomal enzymes and superoxide anions by neutrophils^{5,12,13}. Taken together, IL-8 may be an indispensable link connecting reactive oxygen metabolites with neutrophil infiltration in this model. Furthermore, our data raise the possibility that blocking the action of IL-8 may be beneficial for otherwise intractable reperfusion injuries in human diseases, including myocardial infarction and multiple organ failure. □

6. Larsen, C. G., Anderson, A. O., Appella, E., Oppenheim, J. J. & Matsushima, K. *Science* **243**, 1464–1466 (1989).
7. Schall, T. J., Bacon, K., Toy, K. J. & Goeddel, D. V. *Nature* **347**, 669–671 (1990).
8. Tanaka, Y. et al. *Nature* **361**, 79–82 (1993).
9. Yoshimura, T., Matsushima, K., Oppenheim, J. J. & Leonard, E. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9233–9237 (1987).
10. Detmers, P. A. et al. *J. exp. Med.* **171**, 1155–1162 (1990).
11. Huber, A. R., Kunkel, S. L., Todd, R. T. & Weiss, S. J. *Science* **254**, 99–102 (1991).
12. Thelen, M. et al. *FASEB J.* **2**, 2702–2706 (1988).
13. Schroeder, J.-M., Mrowietz, U., Morita, E. & Christophers, E. *J. Immun.* **139**, 3474–3483 (1987).
14. Furie, M. B., Cramer, E. V., Naprstek, B. L. & Silverstein, S. C. *J. Cell Biol.* **98**, 1033–1041 (1983).
15. Horgan, M. J., Wright, S. D. & Malik, A. B. *Am. J. Physiol.* **259**, L315–L319 (1990).
16. Welbourn, R. et al. *J. Immun.* **145**, 1906–1911 (1990).
17. Hill, J. et al. *Surgery* **112**, 166–172 (1992).
18. Klausner, J. M. et al. *Am. J. Physiol.* **256**, F794–F802 (1989).
19. Egan, J. W. et al. *Prostaglandins* **37**, 597–613 (1989).
20. Zimmerman, B. J., Guillory, D. J., Grisham, M. B., Gaginella, T. S. & Granger, D. N. *Gastroenterology* **99**, 1358–1363 (1990).
21. Lehr, H. A., Guhlman, A., Nolte, D., Keppler, D. & Messmer, K. *J. clin. Invest.* **87**, 2036–2041 (1991).
22. Mentiko, A. P., Kinkel, S. L., Standiford, T. J. & Strieter, R. M. *J. clin. Invest.* **90**, 791–798 (1992).
23. Abe, Y. *Br. Heart J.* (in the press).
24. Lefer, A. M., Johnson, G., Ma, X., Tsao, P. S. & Thomas, G. R. *Br. J. Pharmac.* **103**, 1153–1159 (1991).
25. Gimbrone, M. A. et al. *Science* **246**, 1601–1603 (1989).
26. Lusinskas, F. W. et al. *J. Immun.* **149**, 2163–2171 (1992).
27. Rot, A. *Immun. Today* **13**, 291–294 (1992).
28. Ko, Y.-C. et al. *J. Immunol. Meth.* **149**, 227–235 (1992).
29. Ogata, K., Arakawa, M., Kasahara, T., Nakano, K. & Hiraoka, K. *J. Immunol. Meth.* **65**, 75–82 (1984).
30. DeForge, Fantone, J. C., Kenney, J. S. & Remick, D. G. *J. clin. Invest.* **90**, 2123–2129 (1992).
31. Guesdon, J.-L., Ternyck, T. & Avrameas, S. *J. Histochem. Cytochem.* **27**, 1131–1139 (1979).

ACKNOWLEDGEMENTS. We thank B. W. Nielsen for reviewing this manuscript.

Coupling of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger, Na^+/K^+ pump and sarcoplasmic reticulum in smooth muscle

Edwin D. W. Moore*, Elaine F. Etter*, Kenneth D. Philipson†, Walter A. Carrington*, Kevin E. Fogarty*, Lawrence M. Lifshitz* & Fredric S. Fay*

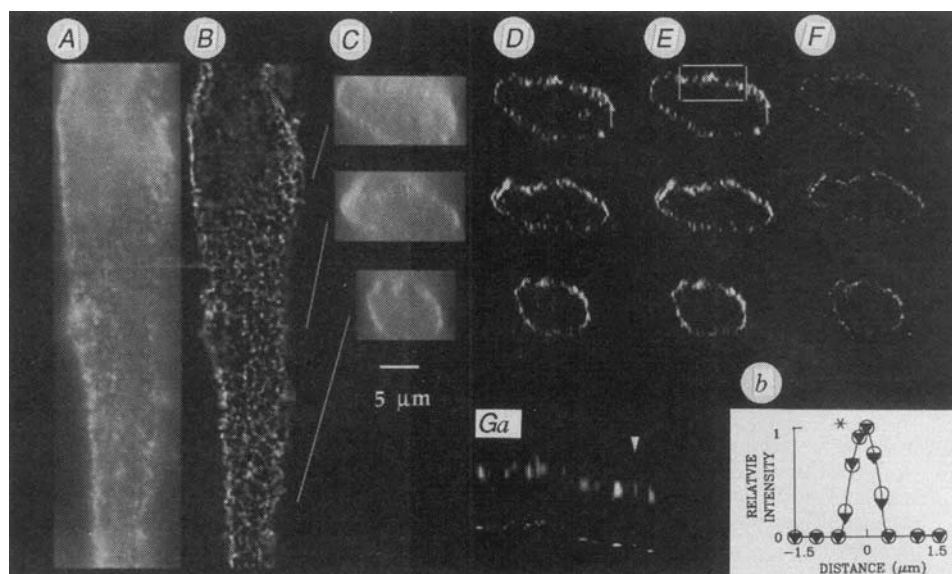
* Program in Molecular Medicine and the Biomedical Imaging Group, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

† Cardiovascular Research Laboratory, Departments of Medicine and Physiology, UCLA School of Medicine, Los Angeles, California 90024, USA

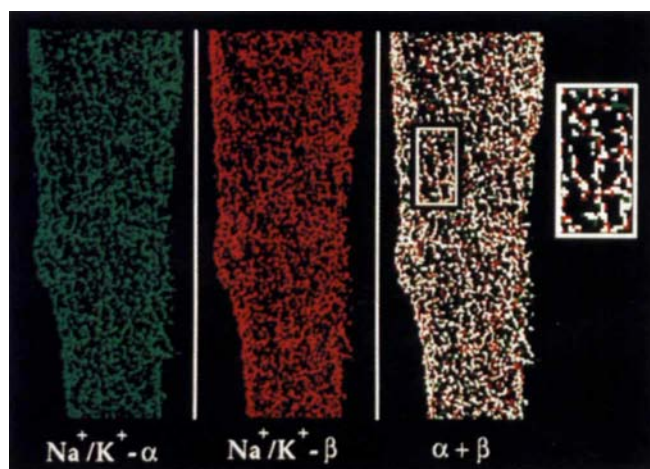
THE $\text{Na}^+/\text{Ca}^{2+}$ exchanger, driven by a transmembrane Na^+ gradient, plays a key role in regulating Ca^{2+} concentration in many cells^{1,2}. Although the exchanger influences Ca^{2+} concentration, its activity in smooth muscle appears to be closely coupled to Ca^{2+} availability from intracellular stores³. This linkage might result if the exchanger were positioned close to Ca^{2+} storage sites within the sarcoplasmic reticulum. To test this hypothesis we have developed methods to assess the relative three-dimensional distribution of proteins involved in Na^+/K^+ pumping, $\text{Na}^+/\text{Ca}^{2+}$ exchange, Ca^{2+} storage within the sarcoplasmic reticulum, and attachment of contractile filaments to the membrane in smooth muscle. Here we report that the $\text{Na}^+/\text{Ca}^{2+}$ exchanger is largely co-distributed with the Na^+/K^+ pump on unique regions of the plasma membrane in register with, and close to, calsequestrin-containing regions of the sarcoplasmic reticulum in sites distinct from the sites where contractile filaments attach to the membrane. This molecular organization suggests that the plasma membrane is divided into at least two functional domains, and appear to provide a mechanism for the strong linkage seen in smooth muscle between Na^+/K^+ pumping and $\text{Na}^+/\text{Ca}^{2+}$ exchange, and between $\text{Na}^+/\text{Ca}^{2+}$ exchange and Ca^{2+} release from the sarcoplasmic reticulum^{4–7}.

Received 8 March; accepted 2 August 1993.

1. Granger, D. N. *Am. J. Physiol.* **255**, H1269–1275 (1988).
2. Welbourn, C. R. B. et al. *Br. J. Surg.* **78**, 651–655 (1991).
3. Romson, J. L. et al. *Circulation* **67**, 1016–1023 (1983).
4. Korthuis, R. J., Grisham, M. B. & Granger, D. N. *Am. J. Physiol.* **254**, H823–H827 (1988).
5. Oppenheim, J. J., Zachariae, C. O. C., Mukaida, N. & Matsushima, K. *A. Rev. Immun.* **9**, 617–648 (1991).



a



b

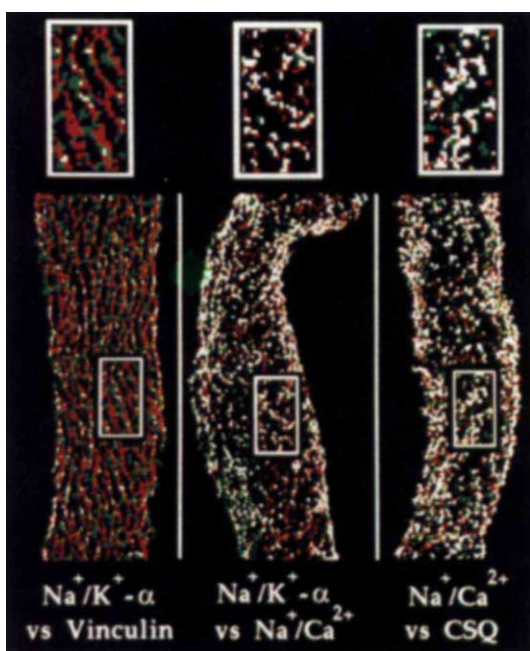


FIG. 2 3D images of the distribution of pairs of proteins involved in Na^+ and/or Ca^{2+} handling in single smooth muscle cells. *a*, Cell labelled for both the α - and β -subunits of the Na^+/K^+ pump. To avoid errors in interpreting the data introduced by superposition of the front and back surfaces, the image of the cell was bisected in the x - y plane; only one half displayed. The α -subunit was labelled with fluorescein and is pseudocoloured green; the β -subunit was labelled with Texas Red and is pseudocoloured red. In the superimposed image, voxels containing signals for both proteins were pseudocoloured white. As the relative stoichiometries of the two proteins are not a simple function of the ratio of intensities of the two fluorophores, the intensity information is not encoded in these images. *b*, Superimposed images of cells dual-labelled for the indicated pairs of proteins. Above each superimposed image in *a* and *b*, the outlined region is displayed at twice the magnification. Left, Na^+/K^+ (α) is green, vinculin red. The overlapped regions of these images are at points where strands of anti-vinculin and strands of anti- Na^+/K^+ pump (α) abutted, reflecting the limits of resolution. By contrast with *a*, strands of anti- α and anti- β - Na^+/K^+ pump are coincident for long stretches across the membrane. Middle, Na^+/K^+ (α) red; $\text{Na}^+/\text{Ca}^{2+}$ exchanger green. Right, $\text{Na}^+/\text{Ca}^{2+}$ exchanger green; calsequestrin red. SR is distributed throughout the cell, but $93.3 \pm 3.6\%$ of the calsequestrin was located within $0.45 \mu\text{m}$ of the plasma membrane. Thus the near-membrane regions of the SR in these cells appears to be functionally differentiated from its deeper elements. Regional ER/SR specialization occurs in vas deferens smooth muscle, which also has a similar peripheral calsequestrin distribution³². Image statistics: *a*, 4,662 surface voxels; Na^+/K^+ pump (α), 23.3%; Na^+/K^+ pump (β), 27.0% and 15.2% contained both; *b* left, 19,762 surface voxels; Na^+/K^+ pump (α), 20.6%; vinculin 26.1%; and 3% contained both; middle, 7,280 surface voxels; Na^+/K^+ pump (α), 22.2%; $\text{Na}^+/\text{Ca}^{2+}$ exchanger, 18.0% and 11.1% contained both; right, 10,602 surface voxels; $\text{Na}^+/\text{Ca}^{2+}$ exchanger 17.2%; calsequestrin, 9.9% and 3.3% contained both. METHODS. Signal in a cell for a given protein at a specific voxel in image *a* was considered to be colocalized with signal in image *b* if it had the same x and y coordinates and a z coordinate no more than 1 voxel away (150 nm). The less stringent criteria for colocalization in z versus x and y was allowed to account for the more limited resolution along the z -axis (Fig. 1G), and the corresponding uncertainty in the true z -position of detected protein. One voxel in *a* was allowed to colocalize with only one voxel in *b*.

In single smooth muscle cells from the stomach of the toad *Bufo marinus*, stimulation of the Na^+/K^+ pump by β -adrenergic agonists enhances $\text{Na}^+/\text{Ca}^{2+}$ exchange and diminishes Ca^{2+} release from the sarcoplasmic reticulum (SR) in response to muscarinic agonists^{8,9}. Although smooth muscle membrane has at least two structurally distinct domains, one rich in dense plaques anchoring actin filaments and the other dominated by caveolae, it is not known where proteins involved in regulating internal Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) and internal Na^+ concentration ($[\text{Na}^+]_i$) are located^{10,11}. We therefore studied the distribution of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger, the Na^+/K^+ ATPase, calsequestrin (a marker for SR Ca^{2+} stores), and vinculin (a marker for contractile filaments anchoring to the plasma membrane) in single, fixed cells.

To obtain three-dimensional (3D) views of cells labelled with antibodies to these molecules, we obtained images at $0.25\text{-}\mu\text{m}$

FIG. 1 Generation of high-resolution 3D images of the distribution of the α -subunit of the Na^+/K^+ pump in a single smooth muscle cell. **A**, Single image plane from a series of optical sections of a smooth muscle cell labelled with an antibody specific for the α -subunit of the Na^+/K^+ pump; the image is dark-current- and background-subtracted, and flat-field-corrected. This, and all subsequent images, are displayed with a gain such that all objects were visible without saturating the film, and the offset was 0. **B**, Same image as **A**, following restoration. **C**, **D**, The cell has been rotated about the x -axis and three cross-sections, 0.45- μm thick, are shown from the indicated regions in the cell: **C**, unrestored; **D**, restored. **E**, Identification of an about-edge annulus, 0.45 μm in either direction from the cell surface. In 38 cells labelled with the five different primary antibodies used here, the annulus contained $95.2 \pm 3.4\%$ (mean $\pm$ s.d.) of the total above-threshold fluorescence; there was no significant difference in annular:total fluorescence for the different antibodies. **F**, Identification of points within the about-edge annulus that are probably the sites giving rise to observed specific signals. **G**, **a**, Magnified views ($\times 1.3$) of region identified in **E** and **F**. **b**, Intensity versus distance from the point of maximum brightness along the z -axis, for a restored image of a 200-nm fluorescent bead (filled triangles) and the cellular region identified by the arrow in **a**. The similarity of these intensity profiles indicates that the molecules on the membrane act as point fluorescent sources; asterisk denotes the point in the image identified by the algorithm described below as the voxel of origin of specific signal. The image restoration procedure significantly reversed blurring introduced by the optics as the full-width at half-maximum intensity of the image of a fluorescein labelled bead before versus after image restoration was 0.32 and 0.23 μm in the x and y directions, and 0.79 and 0.57 μm along the z -axis. **METHODS**. Smooth muscle cells were prepared from *B. marinus* stomach, settled for 90 min on poly-L-lysine-coated coverslips, fixed, permeabilized and labelled with two primary antibodies, as previously described^{24,25}. Secondary antibodies were labelled with fluorescein or

Texas Red (Sigma and Amersham). The primary antibodies were rabbit polyclonals against the α - and β -subunits of the toad kidney Na^+/K^+ pump²⁶, the canine cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger²⁷ and canine calsequestrin²⁸, or mouse monoclonals against chicken gizzard vinculin (Sigma, VIN-11-5). Western blots assessed and confirmed the specificity of these antibodies in this preparation (details available on request). Smooth muscle contains calsequestrin²⁹ and calreticulin³⁰, in toad stomach, calsequestrin is enriched in a fraction also enriched for SR markers. We added 200-nm beads having a broad fluorescent emission (Molecular Probes) to mounting medium containing 1% DABCO as fiducial markers to align images of fluorescein and Texas Red fluorescence. The beads were mounted and imaged identically, providing an empirical measure of the point-spread-function (PSF) of the microscope. All images were acquired with a Nikon $\times 60$, 1.4 NA objective, and a $\times 16$ eyepiece. To separate specific and nonspecific signals, cells labelled with secondary antibody alone were similarly imaged and processed. The mean intensity of cells labelled with both antibodies was at least 6.3 times higher than the mean grey level of cells labelled with secondary antibodies alone. After restoration, we chose a threshold intensity that excluded $>99\%$ of the signal found in the control (no primary antibody) images ($n=5$), and this value was subtracted from images of cells labelled with either the first or second primary antibody, setting any negative-valued pixels to 0. The cell membrane surface was identified in the restored and thresholded image by allowing a function corresponding to a cylindrical surface positioned around the outside of the cell, but not touching it, to deform to the points of maximum intensity in the cell³¹. These surface coordinates were then used to define an annulus, 0.45 μm on either side of this boundary, and voxels outside this region were omitted from further analysis. The remaining annulus data were convolved with the image of a restored 200-nm bead (in effect, a matched filter) to find those points within the annulus that were local maxima along the z -axis. Only the voxels found by this filter were retained for further analysis.

intervals in the digital imaging microscope¹². Although there is almost no antibody labelling either beyond the cell boundaries or in the interior of the cell, fluorescence is seen in these regions because of blurred fluorescence from cellular regions beyond the plane of focus (Fig. 1A, C). To reverse the distortion, we processed this dataset using a constrained deconvolution algorithm based on regularization theory^{13,14}. In restored images (Fig. 1B, D), virtually all the anti- α Na^+/K^+ pump staining is visible, as expected, near the surface of the cell appearing as interconnecting strands that are 0.33 ± 0.02 μm (s.d., $n=57$) wide, running roughly parallel to the long axis of the cell. Because the proteins under investigation are either integral membrane proteins, or are thought to be in near-membrane regions¹⁵, we further processed this and all subsequent images, first to restrict our analysis to regions near the cell surface (Fig. 1E), and then to identify the most likely location of clusters of molecules on that surface (Fig. 1F).

Images of the distribution of pairs of molecules involved in Na^+ and Ca^{2+} handling in single smooth muscle cells are shown in Fig. 2. To visualize the extent to which two molecules in a cell co-distribute, images were superimposed and voxels containing specific signal for only one molecule were either red or green, voxels that contained both were white, and those that had neither remained black. When the 3D images of the distribution of α - and β -subunits of the Na^+/K^+ pumps are superimposed and displayed using this method, most voxels containing specific signal are white, indicating that the pattern of distribution of these molecules is similar (Fig. 2a). Numerical analysis of this and four other similar images reveals that $47.4 \pm 6.3\%$ and $46.0 \pm 6.2\%$ (Table 1) of the signal arising from anti- α - Na^+/K^+ pump and anti- β Na^+/K^+ pump staining, respectively, is present in voxels in which α - and β -subunits coexist.

The functional Na^+/K^+ pump is thought to be a heterodimer^{16,17}, and although we did observe a large and highly significant degree of colocalization, it was less than complete. This might be because in some regions of the cell these subunits do not associate, or alternatively could reflect limits in the extent of overlap detectable. To decide between these possibilities, we used the dual-labelling protocol to tag the β -subunit of the Na^+/K^+ pump twice in the same cell. The extent of overlap of signal resulting

from the first and second exposure to this primary antibody averaged 63%. This was not due to saturation of epitopes by the first antibody exposure (Table 1). Instead, the apparent upper limit observed for the colocalization index is imposed both by the registration of fluorescein and Texas Red images, which because it is done to the nearest pixel, results in a small random residual misregistration of the two fluorescent images and by the immunocytochemical procedure itself which apparently is unable to detect a protein with 100% certainty (Table 1). Although the colocalization index appears to underestimate the actual extent of overlap of two molecular species, it is highly sensitive to small differences in the relative position of molecules in the cell. For molecules clustered on the cell surface like the Na^+/K^+ pump, we calculated that a lateral shift in the pattern of distribution of 50 nm would be reliably detectable ($P \leq 0.05$, $n=5$ cells). Overall, these considerations reveal that the extent of colocalization of the α - and β -subunits would appear to be underestimated by these methods; perhaps more importantly they indicate that these methods are highly sensitive to small shifts in the distribution of two classes of molecules in a cell.

Although there was a large overlap of α - and β -subunits of the Na^+/K^+ pump, there was, in contrast, a striking lack of colocalization of Na^+/K^+ pump (α) and vinculin (Fig. 2b). The extent of overlap of the Na^+/K^+ pump (α) and vinculin signals in this and four similar cells was less than predicted by chance (Table 1). Thus the Na^+/K^+ pump appears to be segregated to a unique region of the plasma membrane, distinct from regions containing sites of attachment of contractile filaments.

To discover whether these patterns reflected a general functional organization of the membrane, and to understand why the activities of the Na^+/K^+ pump and the $\text{Na}^+/\text{Ca}^{2+}$ exchanger appeared to be so closely coupled, we compared the distribution of the Na^+/K^+ pump (α) to that of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger. The pattern of distribution of the exchanger was similar to that the Na^+/K^+ pump, and the proteins appeared highly colocalized (Fig. 2b). These proteins are colocalized to a far greater extent than can be attributed to chance (Table 1). In fact the overlap of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger with the α -subunit of the Na^+/K^+ pump was as large as that recorded when the β -subunit of the

TABLE 1 Numerical analysis of the extent of colocalization of the pairs of indicated proteins

Labelled pair	Observed overlap	Predicted overlap due to chance	Probability observed overlap is due to chance
Na ⁺ /K ⁺ pump: α with β	47.4 $\pm$ 6.3 (5)	25.9	$P < 0.002$
β with α	46.0 $\pm$ 6.2	25.1	
Na ⁺ /K ⁺ (α) with Na ⁺ /Ca ²⁺ exchanger	42.2 $\pm$ 7.3 (5)*	11.6	$P < 0.002$
Na ⁺ /Ca ²⁺ exchanger with Na ⁺ /K ⁺ (α)	62.4 $\pm$ 10.1*	17.2	
Na ⁺ /Ca ²⁺ exchanger with calsequestrin	33.3 $\pm$ 6.3 (4)	12.7	$P < 0.002$
Calsequestrin with Na ⁺ /Ca ²⁺ exchanger	43.3 $\pm$ 7.3	16.4	
Na ⁺ /K ⁺ (α) with vinculin	16.6 $\pm$ 4.6 (5)	25.7	$P < 0.002$
Vinculin with Na ⁺ /K ⁺ (α)	11.7 $\pm$ 3.2	18.1	
Na ⁺ /K ⁺ pump: β (Fl) with β (TxRd)	68.3 $\pm$ 7.0 (4)†		
β (TxRd) with β (Fl)	58.1 $\pm$ 6.8		

The observed overlap is the percentage of voxels (mean $\pm$ s.d.) containing the first listed molecule which also contained the second. The number in parentheses is the number of cells on which this measurement was made. For analysis, a section of the membrane that was roughly normal to the optical axis was segmented from each cell (average 6,600 voxels per cell, or 148.5 μm^2), this was to avoid artefactual overlaps introduced by limited Z resolution. The predicted overlap due to chance is calculated assuming that both proteins of a given pair distribute randomly and independently of each other and is the mean (Np) of a binomial distribution whose variance is $Np(1-p)$, where N is the total number of voxels on the membrane and P is the probability that a voxel contains both proteins. The probability that the observed overlap is the result of chance is the area under this binomial distribution bounded by the observed mean overlap ± 1 s.d.

* These values are significantly different from each other ($P > 0.95$), and may reflect the fact that there was a significantly greater ($P > 0.99$) percentage of the membrane containing signal specific for the α -subunit of the Na⁺/K⁺ pump than signal specific for the Na⁺/Ca²⁺ exchanger (24.9 $\pm$ 5.8% versus 12.9 $\pm$ 4.7% respectively, mean $\pm$ s.d.).

† There was no significant difference in the percentage of the membrane stained with the first (20.7 $\pm$ 5.6%, mean $\pm$ s.d.) versus the second (17.0 $\pm$ 4.9%) exposure to anti-Na⁺/K⁺ pump. The apparent upper limit on the colocalization index is imposed in part by small residual random misalignments of Texas Red and fluorescein image pairs that remains after using fluorescent beads as fiducial markers (Fig. 1 legend). For a typical cell from this series having a colocalization index of β (Fl) and β (TxRd) of 60.6%, residual errors from the registration step are expected to reduce the average colocalization index from 100% to 80.0 $\pm$ 6.6%. We believe that the observed colocalization index is further reduced because the detection of protein by the antibody using these procedures is less than 100%.

pump was labelled twice, suggesting that there are few Na⁺/Ca²⁺ exchangers without a Na⁺/K⁺ pump located within at least 50 nm. The joint clustering to specific regions of the membrane of these two ion-regulating systems would be expected to enhance their coupling, thereby increasing the effectiveness of the Na⁺/K⁺ pump either to influence Na⁺/Ca²⁺ exchange or to maintain Na⁺ gradient in the face of changes in exchanger activity. This molecular organization might also explain the dependence of the SR-releasable Ca²⁺ on the activity of the exchanger, if that store were positioned close to this pump-exchanger complex. We therefore determined the spatial distribution of the Na⁺/Ca²⁺ exchanger relative to that of the SR Ca²⁺ storage protein calsequestrin. Most of the Na⁺/Ca²⁺ exchanger-rich areas of the plasma membrane appear to be closely apposed to and in register with calsequestrin-rich portions of the sarcoplasmic reticulum (Fig. 2b; Table 1).

The molecular organization we have shown here can be related to the ultrastructural specialization of smooth muscle observed with the electron microscope^{10,11,18}. The membrane of these cells has alternating strands of dense plaques and regions dotted with numerous caveolae (caveolar domain). Individual strands often bifurcated as well as merged, and a network of SR was located within 20 nm of the caveolae. Because the dense plaques are rich in α -actinin and vinculin¹⁹, our results indicate that the Na⁺/K⁺ pump and the Na⁺/Ca²⁺ exchanger are in the caveolar domain of the membrane and calsequestrin is in the SR located immediately beneath this domain. The assignment of the pumps and exchangers to the caveolar domain is reinforced by the similarity of the staining pattern (for the Na⁺/K⁺ pumps, strands of beaded staining with an average interbead distance of 470 $\pm$ 68 nm, s.d., $n = 57$) and the spacing of caveolae on the membrane of these cells (about 400 nm; C. R. Scheid, and S. F. Schaeffer, unpublished observations).

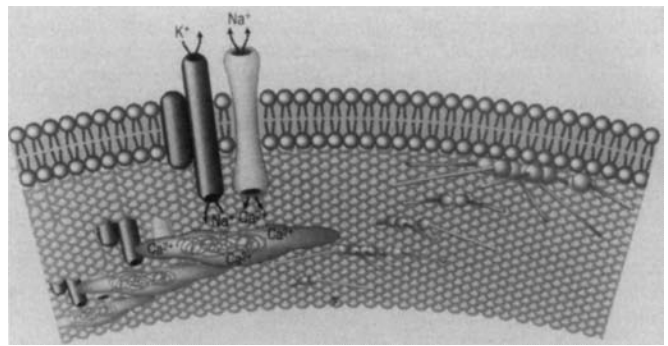


FIG. 3 Model for the distribution of the Na⁺/K⁺ ATPase (α - and β -subunits), the Na⁺/Ca²⁺ exchanger, calsequestrin and vinculin in a smooth muscle cell. Vinculin is located in one macrodomain of the membrane, within the dense plaques, where, in combination with other actin-binding and crosslinking proteins, it anchors actin filaments to the plasma membrane. The Na⁺/K⁺ ATPase and the Na⁺/Ca²⁺ exchanger are in the complementary domain of the membrane enriched in caveolae. Calsequestrin is in the SR, immediately beneath this region of the membrane.

The observed clustering of the Na⁺/K⁺ pump and Na⁺/Ca²⁺ exchanger to regions closely apposed to portions of the SR specialized for Ca²⁺ storage, may have important functional implications for Ca²⁺ homeostasis in these smooth muscle cells. For example, our results can explain the powerful effect of acceleration of the Na⁺/Ca²⁺ exchanger on diminishing SR stores of releasable Ca²⁺. Because of its position, the Na⁺/Ca²⁺ exchanger may have preferred access to Ca²⁺ within the SR across a restricted diffusion space in which ion concentrations may vary significantly from their average cytosolic values^{20,21}. If, as suggested from the structural data, the Na⁺/Ca²⁺ exchanger sees a high Ca²⁺ concentration, as Ca²⁺ diffuses out of or is released from the SR, this would explain the role of the Na⁺/Ca²⁺ exchanger in removing Ca²⁺ from these cells²², despite its apparent low affinity for Ca²⁺ (ref. 23). □

Received 10 May; accepted 21 July 1993.

- Reuter, H. *Nature* **349**, 567–568 (1991).
- Reuter, H., Blaustein, M. P. & Haeusler, G. *Phil. Trans. R. Soc. B* **265**, 87–94 (1973).
- Ashida, T. & Blaustein, M. P. *J. Physiol., Lond.* **392**, 617–635 (1987).
- van Breeman, C. *J. Physiol., Lond.* **272**, 317–329 (1977).
- Bird, G.S.U. et al. *Nature* **352**, 162–165 (1991).
- Brading, A. F. & Sneddon, P. *Br. J. Pharmac.* **70**, 229–240 (1980).
- Stehno-Bittel, L. & Sturek, M. *J. Physiol., Lond.* **451**, 49–78 (1992).
- Scheid, C. R., Honeyman, T. W. & Fay, F. S. *Nature* **277**, 32–36 (1979).
- Williams, D. A. & Fay, F. S. *Am. J. Physiol.* **250**, C779–C791 (1986).
- Somlyo, A. P., Devine, C. E., Somlyo, A. V. & North, S. R. *J. Cell Biol.* **51**, 722–741 (1971).
- Gabella, G. & Blundell, D. *Cell Tiss. Res.* **190**, 255–271 (1978).
- Fay, F. S., Carrington, W. & Fogarty, K. E. *J. Microsc.* **153**, 133–149 (1989).
- Carrington, W. A., Fogarty, K. E. & Fay, F. S. *Modern Cell Biol.* **9**, 53–72 (1990).
- Carrington, W. A. *Soc. Photo-opt. Instr. Engrs Proc.* **1205**, 72–83 (1990).
- Etter, E. F., Jones, L. R., Fogarty, K. E. & Fay, F. S. *J. Cell Biol.* **109**, 169a (1989).
- Fambrough, D. M. *Trends Neurosci.* **11**, 325–328 (1988).
- Horowitz, B. et al. *J. Biol. Chem.* **265**, 4189–4192 (1990).
- Gabella, G. *Phil. Trans. R. Soc. B* **265**, 7–16 (1973).
- Geiger, B. *Cell* **18**, 193–205 (1979).
- LeBlanc, L. & Hume, J. *Science* **248**, 372–376 (1990).
- Niggli, E. & Lederer, W. J. *Science* **250**, 565–568 (1990).
- Fay, F. S., McGeown, J. G., Walsh, J. V. & McCarron, J. G. *Biophys. J.* **61**, A389 (1992).
- Hilgemann, D. W., Nicoll, D. A. & Philipson, K. D. *Nature* **352**, 715–718 (1991).
- Fay, F. S., Hoffman, R., LeClair, S. & Merriam, P. *Meth. Enzym.* **85**, 284–292 (1982).
- Lynch, R. M., Fogarty, K. E. & Fay, F. S. *J. Cell Biol.* **112**, 385–395 (1991).
- Girardet, M., Geering, K., Rossier, B. C., Kraehenbuhl, J. P. & Bron, C. *Biochem. J.* **22**, 2296–2300 (1983).
- Nicoll, D. A., Longoni, S. & Philipson, K. D. *Science* **250**, 562–565 (1990).
- Mahoney, L. & Jones, L. R. *J. Biol. Chem.* **261**, 15257–15265 (1986).
- Wuytack, F., Raeymaekers, L., Verbist, J., Jones, L. R. & Casteels, R. *Biochim. biophys. Acta* **899**, 151–158 (1987).
- Milner, R. E. et al. *J. Biol. Chem.* **266**, 7155–7165 (1991).
- Lifshitz, L. M., Fogarty, K. E., Gauch, J. & Moore, E. D. W. *Soc. Photo-opt. Instr. Engrs Proc.* **1808**, 521–534 (1992).
- Villa, A. et al. *J. Cell Biol.* **121**, 1041–1051 (1993).

ACKNOWLEDGEMENTS. We thank K. Perry and J. Carmichael for technical assistance, J. Collins for software development, D. Bowman for hardware development and support, C. R. Scheid and R. Cooney for assistance with membrane preparations, permission to quote unpublished results, and for discussions, B. Rossier and L. Jones for their gifts of anti-Na⁺/K⁺ pump and anti-calsequestrin, respectively, and M. Michalak in assessing the presence of calreticulin in toad stomach smooth muscle. We acknowledge a postdoctoral fellowship from the Muscular Dystrophy Association and grants from the American Heart Association, Mass. Affiliate (E.D.W.M.), the NIH (F.S.F. and K.D.P.) and the NSF (F.S.F.).

The GTP-binding protein Ran/TC4 is required for protein import into the nucleus

Mary Shannon Moore & Günter Blobel

Laboratory of Cell Biology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Two cytosolic fractions (A and B) from *Xenopus* oocytes are sufficient to support protein import into the nuclei of digitonin-permeabilized cells¹. Fraction A recognizes the nuclear localization sequence (NLS) and binds the import substrate to the nuclear envelope, whereas fraction B mediates the subsequent passage of the bound substrate into the nucleus. Here we report that two interacting components are required for full fraction-B activity, purify one of these components to homogeneity, and show that it is the highly abundant GTP-binding protein Ran (Ras-related nuclear protein)/TC4.

A polypeptide of M_r 25K was purified from *Xenopus* ovarian cytosol on the basis of its ability to stimulate import of an NLS-containing substrate into the nuclei of permeabilized² cells in the

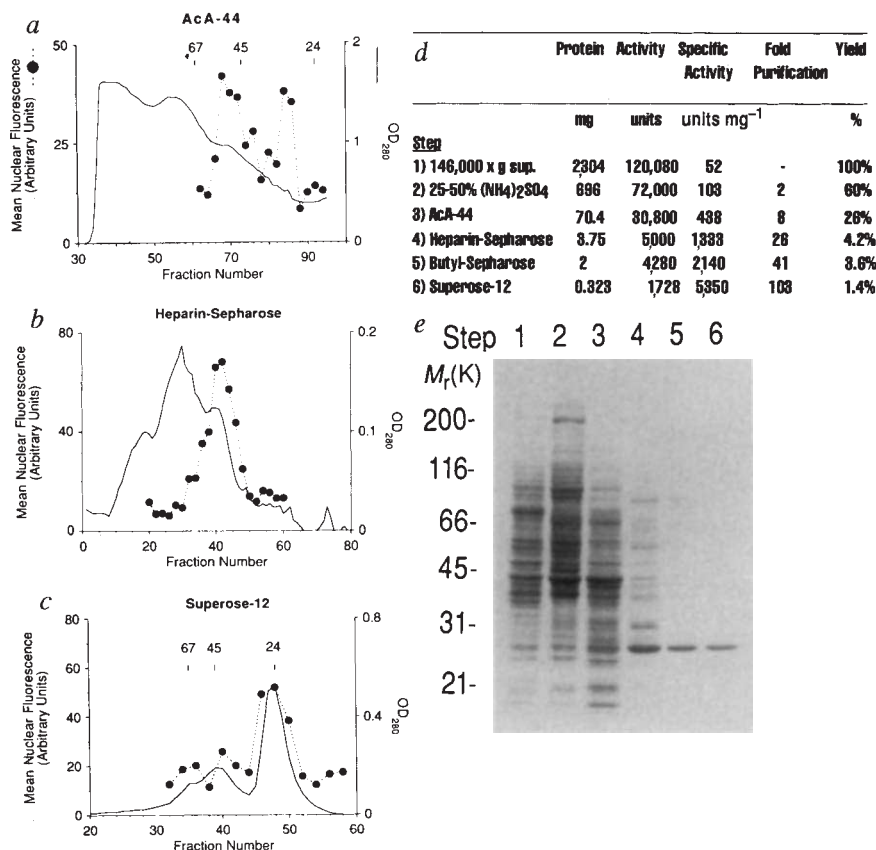
presence of fraction A (Fig. 1). A 103-fold increase in specific activity (Fig. 1d) was sufficient to obtain a pure protein (Fig. 1e). On the first gel filtration column (AcA-44; Fig. 1a) two peaks of B activity were observed, one eluting with an apparent M_r of 50–60K, as previously reported for the active B component¹, and the other eluting at about 30K. On the last purification step (Superose-12; Fig. 1c), however, the activity eluted at an apparent M_r of 25K, corresponding to the size of the single polypeptide present.

Two internal peptide sequences (VATLGVEVHPLVF and FNVWDTAGQEK; single-letter amino-acid code) from the 25K polypeptide cleaved with Lys-C (ref. 3) were identical to predicted amino acids 40–52 and 61–71 of the human protein Ran/TC4 (refs 4–6). Immunoblotting of the purified 25K protein with an anti-Ran/TC4 (human) peptide antiserum confirmed that it was the *Xenopus* homologue of Ran/TC4 (Fig. 2).

Although this protein was isolated on the basis of its ability to support nuclear import (with fraction A), this ability seemed to vary with different preparations and to deteriorate over days at 4 °C during the purification. The purified protein, although able to stimulate the import seen with the A fraction alone, was less effective than crude fraction B¹, indicating that a stimulatory factor had perhaps been lost during the purification. Either *Xenopus* Ran/TC4 or recombinant human Ran/TC4 were therefore incubated with increasing concentrations of crude fraction B (Fig. 3a) in the presence of fraction A. At levels of crude B that

FIG. 1 Purification of the B activity. a–c, Column profiles of steps 3, 4 (gradient elution fractions only) and 6 of the purification (see d). The fractions were assayed for B activity (in the presence of 0.8 mg ml⁻¹ fraction A) as described in the legend to Fig. 3. The elution positions of the calibration standards BSA (67K), ovalbumin (45K), and chymotrypsinogen A (24K) are indicated in a and c. d, Summary of purification; units of B activity are as described¹. e, Coomassie-blue-stained 6–15% SDS-PAGE gel showing the pooled active fractions of each stage of the purification. The lane numbers (e) correspond to the purification steps (d); the following amount of protein was precipitated with trichloroacetic acid and loaded on each lane: 30 µg, lanes 1, 2, 3; 10 µg, lane 4; 5 µg, lane 5; and 2 µg, lane 6. The migration positions of M_r standards are shown at the left.

METHODS. Isolation of *Xenopus* ovarian cytosol in buffer A (20 mM HEPES-KOH, pH 7.3, 2 mM magnesium acetate, 2 mM DTT) and ammonium sulphate precipitation (25–50%) were as described¹. The pellet (~350 mg protein) was resuspended to 15 ml in buffer A containing 80 mM potassium acetate, the suspension was filtered (0.2 µm), and loaded on a 510 ml AcA-44 (Fisher) gel filtration column (2.5 × 104 cm) equilibrated in the same buffer at a flow rate of 30 ml h⁻¹. Fractions (4 ml) were collected and assayed for B activity. The active fractions (66–88 in a) from two separate runs were pooled and diluted with an equal volume of buffer A containing 0.01% Nikkol²⁵. This solution was loaded on a 50 ml heparin-Sepharose (Pharmacia) column (2.5 × 10.2 cm) equilibrated in buffer A containing 40 mM potassium acetate and 0.005% Nikkol at a flow rate of 100 ml h⁻¹. The column was washed with 200 ml of column buffer and eluted with a 500-ml gradient of 40–600 mM potassium acetate in buffer A containing 0.005% Nikkol. Fractions (6 ml) were collected, concentrated (Centricon-10; Amicon), and assayed for B activity. The peak active fractions (37–46 in b) were pooled, brought to 25% saturation with solid ammonium sulphate (13.4 g per 100 ml) and loaded on a 2 ml (1.4 × 1.3 cm) butyl-Sepharose (Pharmacia) column equilibrated in buffer A containing 1.2 M ammonium sulphate at a flow rate of 40 ml h⁻¹. The column was



washed with 20 ml of column buffer and eluted with buffer A containing 100 mM potassium acetate. The protein-containing eluate fractions were pooled (4 ml) and concentrated to 200 µl; 100 µl was injected (in 2 successive runs) on a Pharmacia Superose-12 (FPLC) gel filtration column equilibrated in buffer A containing 100 mM potassium acetate at a flow rate of 0.4 ml min⁻¹. Fractions (200 µl) were collected and assayed. The active fractions (45–51 in c) were pooled, concentrated to 1 mg ml⁻¹, frozen in liquid nitrogen, and stored at -80 °C. The starting material for this purification was the ovarian tissue obtained from 38 frogs (~575 ml settled ovary).

alone sustained little or no import, added *Xenopus* Ran/TC4 produced a marked increase in nuclear import. Ran/TC4 ($50 \mu\text{g ml}^{-1}$), together with 0.5 mg ml^{-1} crude B, sustained import at levels comparable to those achieved with crude B at saturation (2 mg ml^{-1}). Unlike *Xenopus* Ran/TC4, recombinant human Ran/TC4 showed no activity when added with fraction A alone (Fig. 3a; compare the values at $[\text{B}] = 0 \text{ mg ml}^{-1}$) but $50 \mu\text{g ml}^{-1}$ increased import activity slightly relative to crude B. A concentration curve showed that the specific activity of the recombinant Ran/TC4 in this assay was about half that of the *Xenopus* Ran/TC4 (Fig. 3b), perhaps reflecting species differences or, more likely, differences in the purification procedures. Another small GTP-binding protein (p21^{H-ras}) had no effect in the assay, indicating a specific requirement for Ran/TC4.

Fluorescence micrographs of the import assay done in the presence of the various factors show that fraction A alone (Fig. 4A, d) gives a characteristic peripheral nuclear staining and a small amount of import as previously described¹, whereas either Ran/TC4 (Fig. 4A, c) or a low concentration of crude B alone (Fig. 4A, b) had no effect. Fraction A plus Ran/TC4 increased import (Fig. 4A, f) but a high level of import required the combination of A, Ran/TC4, and a low concentration of crude B (Fig. 4A, g). Import was abolished by wheat-germ agglutinin (WGA) (Fig. 4A, h), a known inhibitor of nuclear pore complex function^{7,8}, confirming that the observed increases in intranuclear fluorescence were due to import and not to a disruption of the nuclear envelope.

To investigate whether Ran/TC4 has to be in the 'active', GTP-bound state to mediate nuclear import, GTP, GDP- β S, or GMP-PNP were added to import reactions containing fraction

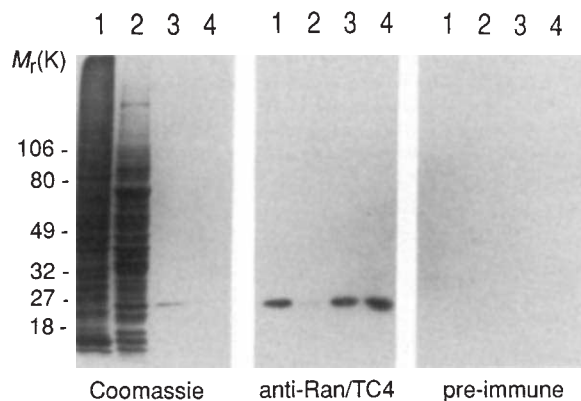


FIG. 2 The purified 25K protein is Ran/TC4. Lanes are: 1, $50 \mu\text{g}$ total HeLa cell lysate; 2, $50 \mu\text{g}$ *Xenopus* ovarian cytosol ($146,000\text{g}$ supernatant); 3, 500 ng *Xenopus* protein isolated in the purification shown in Fig. 1; and 4, 250 ng of human recombinant Ran/TC4. Proteins were separated by SDS-PAGE and either stained with Coomassie blue (left panel) or transferred to nitrocellulose and probed with an antiserum to a human Ran/TC4 peptide (middle panel) or preimmune serum (right panel). Shown at the left are the migration positions of M_r standards. METHODS. The peptide (C)QYEHDLVAQTT (corresponding to amino acids 196–207 of the human Ran/TC4 sequence⁵ plus an amino-terminal cysteine for coupling purposes) was synthesized, coupled to keyhole limpet haemocyanin with MBS (Pierce), and used to produce antibodies in rabbits. Protein samples were subjected to SDS-PAGE and transferred to nitrocellulose. Immunoblotting was as described⁵, except that rabbit sera were diluted 1:500 and the primary incubation was done overnight at 4°C . To obtain the HeLa cell lysate, washed and pelleted HeLa cells were incubated with 1% SDS, boiled (5 min), sonicated and microfuged. The resulting supernatant was used for loading. Human recombinant Ran/TC4 was a gift from E. Coutavas, M. Rush and P. D'Eustachio (NYU Medical Center, New York). The methods for the expression and purification of recombinant Ran/TC4 from *Escherichia coli* will be presented elsewhere (E. Coutavas, M. Ren, J. D. Oppenheim, P. D'Eustachio and M. G. Rush, manuscript in preparation).

A and a low concentration of crude B either with or without Ran/TC4. None of these nucleotides affected the rim staining and the small amount of import obtained with fraction A and crude B alone (Fig. 4B, a, c, e), but both GDP- β S and GMP-PNP inhibited the increased import mediated by Ran/TC4. In the presence of GDP- β S, Ran/TC4 had no effect (Fig. 4B, c, d), whereas in the presence of the non-hydrolysable GTP analogue GMP-PNP the rim staining and small amount of import obtained in the absence of Ran/TC4 were reversed (or pre-

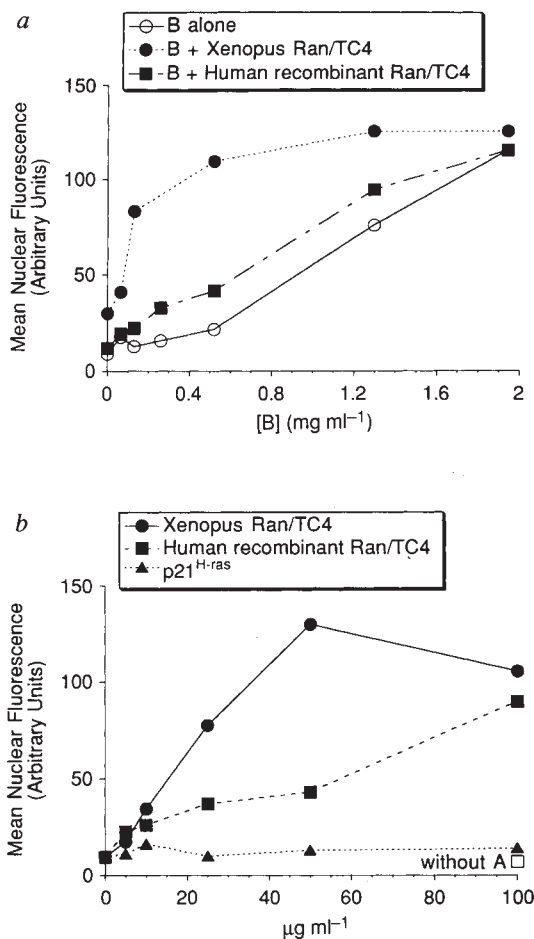


FIG. 3 Ran/TC4 requires a stimulatory factor for full activity. a, Increasing concentrations of crude fraction B were assayed (in the presence of fraction A) for import activity either alone (open circles), with $50 \mu\text{g ml}^{-1}$ purified *Xenopus* Ran/TC4 (solid circles), or with $50 \mu\text{g ml}^{-1}$ human recombinant Ran/TC4 (squares). b, Crude fraction B (0.4 mg ml^{-1}) was incubated with fraction A and increasing concentrations of either *Xenopus* Ran/TC4 (circles), human recombinant Ran/TC4 (solid squares) or recombinant p21^{H-ras} (triangles). The empty square shows the value obtained with human recombinant Ran/TC4 assayed with crude fraction B but without fraction A.

METHODS. Digitonin-permeabilized buffalo rat liver (BRL) cells were incubated in the standard assay mixture (1 mg ml^{-1} BSA, 1 mM ATP, 5 mM phosphocreatine, and 20 U ml^{-1} creatine phosphokinase in buffer A containing 110 mM potassium acetate and 1 mM EGTA) for 15 min at room temperature before washing and fixation^{1,2}. The import substrate (rhodamine-labelled human serum albumin coupled to peptide containing the nuclear localization sequence of the SV40 T antigen)¹ was added at $5 \mu\text{g ml}^{-1}$. *Xenopus* fraction A¹ was added at a concentration of 0.8 mg ml^{-1} to all samples in a and b except where noted. The crude fraction B was obtained by ammonium sulphate precipitation (60%) from the DE-52 flow-through fractions obtained during the separation of fractions A and B as described¹. Quantification was as described¹ and between 10 and 30 nuclei were scanned per point. The p21^{H-ras} (ref. 26) was a gift from M. Brown and J. Goldstein.

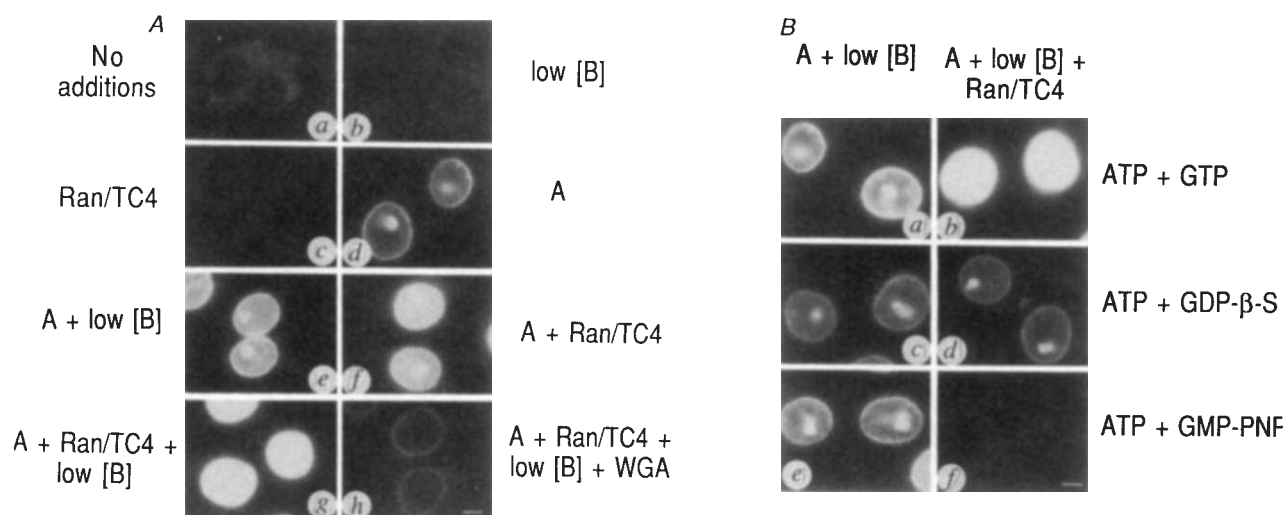


FIG. 4 Fluorescence microscopy of the import assay under various conditions. A, Import conditions were as described in the legend to Fig. 3. Where indicated, the following components were incubated together: fraction A (0.8 mg ml^{-1}), low [B] (crude fraction B at 0.4 mg ml^{-1}), and purified *Xenopus* Ran/TC4 ($50 \mu\text{g ml}^{-1}$). Wheat-germ agglutinin (WGA; Vector) 0.5 mg ml^{-1} . Scale bar, $10 \mu\text{m}$. B, Fraction A, low [B], and

Xenopus Ran/TC4 were the same as in A. The ATP regenerating system was deleted in this assay and the incubation included 1 mM ATP plus either 1 mM GTP, 1 mM GDP- β S, or 1 mM GMP-PNP (all from Boehringer-Mannheim). All of the micrographs in each panel were from the same experiment and were exposed and printed for the same length of time. Scale bar, $10 \mu\text{m}$.

vented) (Fig. 4B, e, f), giving a staining pattern identical to samples that had never received fraction A (Fig. 4A, a-c). GTP hydrolysis is thus required for nuclear import, but it is still unclear whether it is performed by Ran/TC4 or another GTPase.

Ran/TC4 represents 0.36% of the total protein in HeLa cells⁶, of which about 80% is nuclear and is not released on permeabilization of cells with digitonin⁵. This nuclear Ran/TC4 is apparently unable to substitute for the cytoplasmic pool in the import assay, although the small amount of import seen with the A fraction alone (ref. 1; and Fig. 4A) might be due to this nuclear pool.

Ran/TC4 forms a complex with an abundant chromatin-binding protein called RCC1 (regulator of chromosome condensation) which catalyses nucleotide exchange on Ran/TC4 *in vitro*^{6,9,10}. RCC1 affects numerous nuclear events, including DNA replication, the cell cycle, maintenance of nuclear structure, the mating response (in yeast), and RNA transcription, processing, and export^{5,11,21} (reviewed in ref. 22). These pleiotropic effects might reflect a role for RCC1 in maintaining the active GTP-bound form of Ran/TC4, and the stimulatory activity that we have observed in the crude B fraction might be

due to a cytoplasmic protein serving the same function. If some of the purified Ran/TC4 was bound to GTP, this might explain its limited activity without crude B (unlike recombinant Ran/TC4). The B fraction contains no Hsc70 (refs 23 and 24, and data not shown) and the $50\text{--}60\text{K}$ peak of B activity seen in Fig. 1a is too small to be the RCC1-Ran/TC4 complex⁶; it may represent a complex between the putative exchange protein and Ran/TC4.

The precise function of Ran/TC4 in nuclear import remains unclear. It may shuttle between the nucleus and the cytoplasm in association with carriers that function as receptors for cargo to be imported into or exported from the nucleus, with GTP-Ran/TC4 facilitating the cargo-carrier association and GDP-Ran/TC4 facilitating dissociation. This cargo would thus consist of proteins or ribonucleoproteins (and perhaps deoxyribonucleoproteins in the case of viruses) and the A fraction would contain at least one of the carriers. Cargo-carrier dissociation could then be controlled by localized GTPase-activating proteins (GAPs), whereas association would be mediated by proteins like RCC1. Such a mechanism would allow deposition of import substrate by localized GAPs and acquisition of export substrate at suitable points along the import/export pathway. □

Received 18 March; accepted 23 August 1993.

- Moore, M. S. & Blobel, G. *Cell* **69**, 939-950 (1992).
- Adam, S. A., Sterne-Marr, R. & Gerace, L. J. *Cell Biol.* **111**, 807-816 (1990).
- Fernandez, J. M., DeMott, D., Atherton, D. & Mische, S. M. *Analyt. Biochem.* **201**, 255-264 (1992).
- Drivas, G. T., Shih, A., Coutavas, E., Rush, M. G. & D'Eustachio, P. *Molec. cell. Biol.* **10**, 1793-1798 (1990).
- Ren, M., Drivas, G., D'Eustachio, P. & Rush, M. G. *J. Cell Biol.* **120**, 313-323 (1993).
- Bischoff, F. R. & Ponstingl, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10830-10834 (1991).
- Finlay, D. R., Newmeyer, D. D., Price, T. M. & Forbes, D. J. *J. Cell Biol.* **104**, 189-200 (1987).
- Yoneda, Y., Imamoto-Sonobe, N., Yamaizumi, M. & Uchida, T. *Exp. Cell Res.* **173**, 586-595 (1987).
- Bischoff, F. R. & Ponstingl, H. *Nature* **354**, 80-82 (1991).
- Bischoff, F. R., Maier, G., Tilz, G. & Ponstingl, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8617-8621 (1990).
- Nishimoto, T., Eilen, E. & Basilico, C. *Cell* **15**, 475-483 (1978).
- Ohtsubo, M., Okazaki, H. & Nishimoto, T. *J. Cell Biol.* **109**, 1389-1397 (1989).
- Uchida, S. *et al. Molec. cell. Biol.* **10**, 577-584 (1990).
- Matsumoto, T. & Beach, D. *Cell* **66**, 347-360 (1991).

- Dasso, M., Nishitani, H., Kornbluth, S., Nishimoto, T. & Newport, J. W. *Molec. cell. Biol.* **12**, 3337-3345 (1992).
- Aebi, M., Clark, M. W., Vijayraghavan, U. & Abelson, J. *Molec. gen. Genet.* **224**, 72-80 (1990).
- Forrester, W., Stutz, F., Rosbash, M. & Wickens, M. *Genes Dev.* **2**, 754-765 (1992).
- Amberg, D. C., Fleischmann, M., Stagljar, I., Cole, C. N. & Aebi, M. *EMBO J.* **12**, 233-241 (1993).
- Ohtsubo, M. *et al. EMBO J.* **10**, 1265-1273 (1991).
- Vijayraghavan, U., Company, M. & Abelson, J. *Genes Dev.* **3**, 1206-1216 (1989).
- Clark, K. L. & Sprague, G. F. Jr. *Molec. cell. Biol.* **9**, 2682-2694 (1989).
- Dasso, M. *Trends bio. Sci.* **18**, 96-101 (1993).
- Shi, Y. & Thomas, J. O. *Molec. cell. Biol.* **12**, 2186-2192 (1992).
- Imamoto, N. *et al. J. Cell Biol.* **119**, 1047-1061 (1992).
- Connolly, T. & Gilmore, R. J. *Cell Biol.* **103**, 2253-2261 (1986).
- Reiss, Y., Goldstein, J. L., Seabra, M. C., Casey, P. J. & Brown, M. S. *Cell* **62**, 81-88 (1990).

ACKNOWLEDGEMENTS. We thank E. Coutavas, M. Rush, and P. D'Eustachio for recombinant Ran/TC4, M. Brown and J. Goldstein for the p21^{H-ras}, the staff of the Rockefeller Biopolymer Facility for protein sequence analysis, and S. Wente, P. Bernstein, C. Nicchitta, S. Smith and T. Meier for a critical reading of the manuscript and many helpful discussions.

ATP-induced protein-Hsp70 complex dissociation requires K^+ but not ATP hydrolysis

Daniel R. Palleros*, Katherine L. Reid*, Li Shi*, William J. Welch† & Anthony L. Fink*‡

* Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064, USA

† Departments of Medicine and Physiology, University of California, San Francisco, California 94143, USA

THE molecular chaperone proteins, particularly Hsp60 and Hsp70, have been implicated in essential cell functions under both normal and stress conditions (reviewed in refs 1–5). Members of the family of heat-shock proteins of M_r 70K, Hsp70, bind to unfolded proteins and short peptides^{6–13}. Addition of Mg-ATP results in the dissociation of the substrate polypeptides from the chaperone^{7–11}, but as ATP- γ S (an ATP analogue that is only slowly hydrolysable) cannot substitute for ATP in this reaction^{7,9,11}, it has been concluded that ATP hydrolysis is necessary to dissociate Hsp70-substrate protein complexes. By independently measuring the rates of ATP hydrolysis and substrate protein dissociation, we show here that Mg-ATP binding but not Mg-ATP hydrolysis is essential for substrate dissociation. We also show that there is an absolute requirement for K^+ for the effect of Mg-ATP: only the combination of K^+ and Mg-ATP will cause the conformational change in Hsp70 that is necessary for substrate dissociation. Moreover, in the absence of K^+ , Mg-ATP favours complex formation. We consider these results in terms of a G-protein-like model.

We studied the interactions of DnaK from *Escherichia coli*, its mutant T199A (which has impaired ATPase activity¹⁴), bovine brain Hsp73 and human Hsp72 (ref. 15) with reduced carboxymethylated α -lactalbumin (RCMLA), a permanently unfolded form of α -lactalbumin⁷, and a thermally unstable mutant of staphylococcal nuclease, NCA-SNase^{7,16}. Figure 1 shows the traces from size-exclusion HPLC (SEC-HPLC) of DnaK and DnaK-RCMLA complex; analysis of the isolated complex by SDS-PAGE indicated a 1:1 stoichiometry (SEC-HPLC results are summarized in Table 1). When DnaK was incubated for 10 min at 37 °C in the presence of an excess of RCMLA or NCA-SNase, and the mixture analysed by SEC-HPLC using a mobile-phase buffer containing 200 mM KCl, which is the intracellular concentration of K^+ in *E. coli*, about 48% of the Hsp70 was found associated with the substrate protein (formation of 100% complex requires longer incubation or/and higher substrate concentration). Addition of Mg-ATP to the reaction mixture and immediate injection into the HPLC resulted in essentially complete dissociation of the complex. Addition of Mg-ATP- γ S did not result in complex dissociation. Addition of Mg-ADP (or Mg-AMP; data not shown) did not dissociate the complex; on the contrary, there was a slight increase in the proportion of complex. Similar results were obtained with the other Hsp70s and also with NCA-SNase.

We did the same experiments using a mobile-phase buffer containing Na^+ instead of K^+ and obtained very different results: when Mg-ATP was added to the Hsp70-protein complex in the absence of K^+ immediately before injection into HPLC, the complex did not dissociate; but when an excess of KCl was added with Mg-ATP and the mixture injected into HPLC with NaCl in the mobile phase, there was no significant change in the amount of complex (entry 19 shown in Table 1). This indicated that the complex, presumably dissociated by K^+ /Mg-ATP, had reformed during injection into HPLC, as K^+ was retarded in the HPLC column. In the absence of KCl, Mg-ATP favours the Hsp70-protein interaction, as the results shown

in Table 1 indicate. When a mixture of DnaK, RCMLA and Mg-ATP (no K^+) was immediately injected into HPLC (with NaCl), about 60% of the complex was detected (entry 28). Under similar conditions, but using KCl in the HPLC buffer, less than 2% complex remained (entry 14). Further confirmation that complex formation is faster in the presence of Mg-ATP without K^+ came from adding DnaK, RCMLA and Mg-ATP successively to the HPLC injection loop, then injecting at once so that mixing took place inside the HPLC; under these conditions, 40% complex was detected. When 200 mM KCl was present in the HPLC buffer, less than 2% complex formed; the complex between substrate protein and Hsp70 does not reform after separation of ATP in the HPLC column because under these conditions (22 °C, 200 mM KCl) complex formation is slow^{6,7}. We confirmed the effects of Na^+ and K^+ on the nucleotide-induced protein binding and dissociation in experiments using 50 μ M Mg-ATP or Mg-ADP in the SEC-HPLC eluting buffers.

The different effects exerted by Na^+ and K^+ in the presence of Mg-ATP are manifested in a conformational change of Hsp70. Figure 2 shows the tryptophan fluorescence emission spectra of

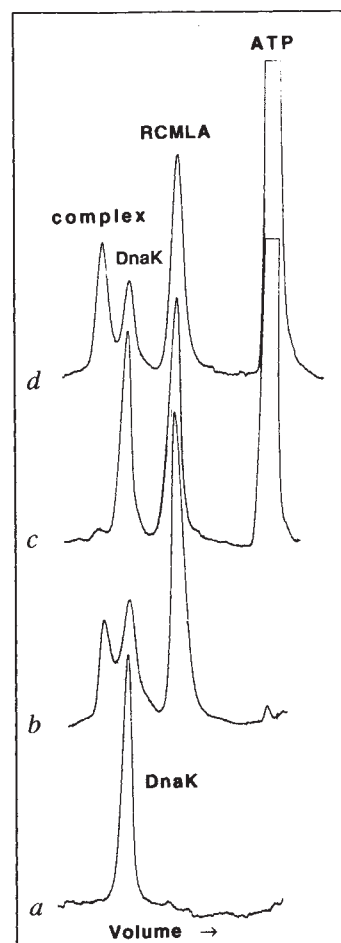


FIG. 1 SEC-HPLC traces of DnaK and DnaK-RCMLA complex. HPLC analysis were carried out as already described using a 20- μ l injection loop⁶. DnaK was prepared as previously described⁶. For a–c, the mobile phase was 20 mM sodium phosphate, 200 mM KCl, pH 6.5. a, DnaK, 4.5 μ M in 100 mM Tris-HCl, pH 7.0; b, DnaK was incubated with RCMLA for 10 min at 37 °C and injected (entry 1 in Table 1); c, DnaK was incubated with RCMLA for 10 min at 37 °C, Mg-ATP was added and injected (entry 2, Table 1); d, DnaK was mixed with RCMLA and Mg-ATP and immediately injected into the HPLC equilibrated with 20 mM sodium phosphate, 200 mM sodium chloride, pH 6.5 (entry 28 in Table 1). Elution volumes (ml) are: 9.5 (void volume); 12.4 (DnaK-RCMLA complex); 13.9 (DnaK); 16.5 (RCMLA); 21.8 (ATP); flow rate, 1 ml min⁻¹.

‡ To whom correspondence should be addressed.

TABLE 1 Binding and dissociation of unfolded proteins to Hsp70 followed by SEC-HPLC

HPLC buffer: 200 mM KCl; 20 mM sodium phosphate; pH 6.5

	% Complex
1. DnaK + RCMLA	48
2. DnaK + RCMLA + ATP	5
3. DnaK + RCMLA + ADP	66
4. DnaK + RCMLA + ATP- γ S	37
5. DnaK + NCA-SNase	48
6. DnaK + NCA-SNase + ATP	5
7. DnaK + NCA-SNase + ADP	63
8. DnaK + NCA-SNase + ATP- γ S	40
9. DnaK T199A + RCMLA	64
10. DnaK T199A + RCMLA + ATP	8
11. Hsp72 + RCMLA	70
12. Hsp72 + RCMLA + ATP	27
13. DnaK + RCMLA*	2
14. DnaK + RCMLA + ATP*	2
15. DnaK + RCMLA + ADP*	20
16. DnaK + RCMLA + ATP loop†	2

HPLC buffer: 200 mM NaCl; 20 mM sodium phosphate; pH 6.5

17. DnaK + RCMLA	55
18. DnaK + RCMLA + ATP	53
19. DnaK + RCMLA + ATP/KCl	53
20. DnaK + RCMLA + ADP	58
21. DnaK + NCA-SNase	41
22. DnaK + NCA-SNase + ATP	37
23. Hsp72 + RCMLA	50
24. Hsp72 + RCMLA + ATP	47
25. Hsp72 + NCA-SNase	41
26. Hsp72 + NCA-SNase + ATP	40
27. DnaK + RCMLA*	19
28. DnaK + RCMLA + ATP*	60
29. DnaK + RCMLA + ADP*	15
30. DnaK + RCMLA + ATP/KCl*	47
31. DnaK + RCMLA + ATP- γ S*	21
32. DnaK + RCMLA loop†	5
33. DnaK + RCMLA + ATP loop†	39
34. DnaK T199A + RCMLA*	5
35. DnaK T199A + RCMLA + ATP*	62

Column on right refers to per cent of DnaK in complex, as determined by peak areas. DnaK T199A was prepared as described¹⁴; Hsp72 was isolated as reported elsewhere¹⁵. HPLC mobile phase was either 20 mM sodium phosphate, 200 mM KCl, pH 6.5 (entries 1–16), or 20 mM sodium phosphate, 200 mM NaCl, pH 6.5 (entries 17–35). Stock solutions: DnaK stock solution was 19.4 μ M in 18 mM Tris-HCl, 45 mM NaCl, 5 mM β -mercaptoethanol, 10% glycerol, pH 7.5; DnaK T199A stock solution was 4.32 μ M in 19 mM Tris-HCl, 1.25 mM HEPES, 50 μ M EDTA, 250 μ M β -mercaptoethanol, 0.5% glycerol, 20 mM KCl, pH 7.5; Hsp72 stock solution was 4.6 μ M in 20 mM Tris-HCl, 100 mM NaCl, 10 mM KCl, 5 mM β -mercaptoethanol, pH 7.4; RCMLA stock solution was 327 μ M in HCl, pH 2; NCA-SNase stock solution was 368 μ M in 20 mM Tris-HCl, pH 7.5; ATP stock solution was 10 mM disodium ATP, 10 mM MgCl_2 , 20 mM Tris-HCl, pH 6.5; ADP stock solution was 9.1 mM monosodium ADP, 10 mM MgCl_2 , 100 mM Tris-HCl, pH 6.5; ATP- γ S stock solution was 50 mM tetralithium ATP- γ S, 50 mM MgCl_2 , 20 mM sodium phosphate, 200 mM KCl, pH 6.5. Reaction: Entries 1–12 and 17–26: Hsp70 stock solution was mixed with substrate protein stock solution, a variable volume (as indicated below) of 20 mM Tris-HCl, pH 7.5, was added and the mixture was incubated at 37 °C for 10 min; nucleotide stock solution was added and the mixture injected without delay into the HPLC; entries 11 and 12: incubation was for 30 min instead of 10 min; entry 12: ATP stock solution was in 20 mM sodium phosphate instead of Tris-HCl; phosphate has inhibitory effects on complex dissociation (D.R.P. and A.L.F., unpublished results). Entries 13–15, 27–31 and 34–35: Hsp70 stock solution was mixed with RCMLA and the nucleotide (if appropriate) stock solutions and immediately injected into the HPLC; entries 16, 32 and 33: DnaK, RCMLA and ATP stock solutions (if appropriate) were successively added into the HPLC injection loop and injected at once; entries 1–3, 5–7, 13–22, 27–30, 32 and 33: 7 μ l DnaK stock solution + 2 μ l RCMLA or NCA-SNase stock solution + 21 μ l Tris-HCl (20 mM, pH 7.5) + 6 μ l of nucleotide stock solution (if indicated); entries 19 and 30 as indicated above + 10 μ l 3 M KCl in water added after nucleotide addition; entries 4, 8 and 31: 7 μ l DnaK stock solution + 2 μ l of RCMLA or NCA-SNase stock solutions + 21 μ l Tris-HCl (20 mM, pH 7.0) + 1 μ l ATP- γ S stock solution entries 9–10, 34 and 35: 28 μ l DnaK stock solution + 2 μ l of RCMLA stock solution + 6 μ l of ATP stock solution if indicated; entries 11, 12, 23–26: 28 μ l Hsp72 stock solution + 2 μ l RCMLA or NCA-SNase stock solutions + 6 μ l nucleotide stock solution (if indicated).

* Immediate injection.

† Mixed in injection loop.

observed in the presence of NaCl or KCl. Addition of Mg-ATP (no K^+) resulted in a decrease in the quantum yield and no significant change in λ_{max} . The presence of 200 mM NaCl with Mg-ATP caused a blue shift to 328 nm. A more significant effect was seen with 200 mM KCl and Mg-ATP, with λ_{max} at 324 nm. Similar results were obtained with DnaK T199A (not shown). Neither Mg-ADP nor Mg-ATP- γ S (with or without K^+), caused a blue shift (Fig. 2b). As ADP, ATP and ATP- γ S all bind to Hsp70 with comparable affinity in the μ M range of concentration^{7,18}, it is likely that the conformational change caused by K^+ /MgATP is the result of a very specific interaction.

The rate of ATP hydrolysis (Fig. 3) in the presence of RCMLA by wild-type DnaK and DnaK T199A at 37 °C in the presence of 200 mM K^+ was 0.18 ± 0.02 and $0.011 \pm 0.002 \text{ min}^{-1}$, respectively. Also, there was no fast burst of ATP hydrolysis corresponding to a single turnover of complex when the complex between DnaK (0.5 or 2.4 μ M) and RCMLA formed and Mg-ATP was added; the extrapolated amount of ADP produced at time zero was $<0.1 \mu$ M for both DnaK concentrations. The effect of KCl was also evident on both the intrinsic and RCMLA-stimulated ATPase activity of wild-type DnaK; lack of K^+ resulted in a twofold decrease in the rate of ATP hydrolysis.

The hydrolysis of a stoichiometric amount of ATP requires at least 5 min for wild-type DnaK and 88 min for its mutant at 37 °C (and therefore considerably longer at 22 °C). In contrast, according to our HPLC results, complex dissociation must be completed in less than 12 min (the elution time for the DnaK-RCMLA complex); in fact the width and shape of the free DnaK peak (Fig. 1c) indicate that the substrate must have dissociated

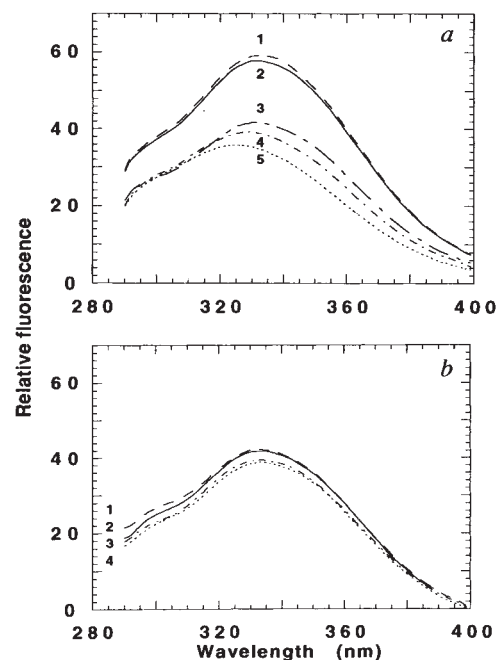


FIG. 2 Fluorescence spectra of DnaK at 20 °C. Fluorescence spectra were determined as before, using a 1-ml quartz cuvette⁶. Solutions were prepared by mixing 80 μ l DnaK stock solution (8.7 μ M in 100 mM Tris-HCl, pH 7.0), aliquots of nucleotide stock solution (details in Table 1 legend), NaCl or KCl solutions (3 M in water) and Tris-HCl buffer (100 mM, pH 7.0) to make up a final volume of 870 μ l and the indicated final concentrations of nucleotides and added salts. Spectra in a, (1) 200 mM KCl; (2) no nucleotides or salts added; (3) 1 mM Mg-ATP; (4) 1 mM Mg-ATP + 200 mM NaCl; (5) 1 mM Mg-ATP + 200 mM KCl. Spectra in b, (1) 1 mM Mg-ATP- γ S; (2) 1 mM Mg-ADP; (3) 1 mM Mg-ATP- γ S + 200 mM KCl; (4) 1 mM Mg-ADP + 200 mM KCl.

DnaK; the single tryptophan residue is located in the N-terminal domain¹⁷. In the absence of nucleotides, DnaK has a λ_{max} at 331 nm; no significant change in λ_{max} or quantum yield is

within seconds of injection. Corroboration that ATP hydrolysis did not occur during substrate dissociation was obtained in a SEC-HPLC experiment in which α - and γ - ^{32}P -labelled ATP were added to DnaK-RCMLA complex (with K^+) before injection. In each case essentially the same concentration of radioisotope was associated with DnaK, indicating that the ADP formed was negligible although dissociation of the complex was complete. Thus, at least *in vitro*, substrate dissociation precedes ATP hydrolysis and is brought about by the specific binding of ATP in the presence of K^+ .

Our results can be explained by the cycle shown in Fig. 3b, in which the concentrations of K^+ and nucleotides are physiological, and h and h* represent high-affinity and low-affinity (induced by binding ATP) conformations of Hsp70 for substrate, respectively. We have previously shown that under physiological conditions Hsp70 will exist as a binary complex with nucleotide⁷. This cycle is remarkably similar to that of G proteins (GTPases), in which binding and hydrolysis of GTP

drive transitions between two conformational states, GDP-bound and GTP-bound¹⁹. Whereas for G proteins it is the GTP-bound form that is active, in the case of Hsp70 it is the ADP-bound form, in the sense of having a high binding affinity for unfolded substrates⁷. We believe that the cofactor GrpE and its homologues^{12,20} will function as nucleotide-exchange proteins in catalysing the exchange of ATP for ADP (reaction 2 in Fig. 3b). It is likely that DnaJ and its homologues^{12,20} will interact (reaction 4) to increase the ATP hydrolysis rate. □

Received 18 March; accepted 21 July 1993.

1. Ellis, R. J. *Seminars Cell Biol.* **1**, 1–9 (1991).
2. Hartl, F. U., Martin, J. & Neupert, W. A. *Rev. Biophys. biomolec. Struct.* **21**, 293–322 (1992).
3. Gething, M.-J. & Sambrook, J. *Nature* **355**, 33–45 (1992).
4. Ang, D., Liberek, K., Skowrya, D., Zylitz, M. & Georgopoulos, C. *J. biol. Chem.* **266**, 24233–24236 (1991).
5. Hightower, L. E. *Cell* **66**, 191–197 (1991).
6. Palleros, D. R., Reid, K. L., McCarty, J. S., Walker, G. C. & Fink, A. L. *J. biol. Chem.* **267**, 5279–5285 (1992).
7. Palleros, D. R., Welch, W. J. & Fink, A. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5719–5723 (1991).
8. Beckmann, R. P., Mizzen, L. A., & Welch, W. J. *Science* **248**, 850–854 (1990).
9. Flynn, G. C., Chappell, T. G. & Rothman, J. E. *Science* **245**, 385–390 (1989).
10. Landry, S. J., Jordan, R., McMacken, R. & Gierasch, L. M. *Nature* **355**, 455–457 (1992).
11. Liberek, K., Skowrya, D., Zylitz, M., Johnson, C. & Georgopoulos, C. *J. biol. Chem.* **266**, 14491–14496 (1991).
12. Langer, T. *et al. Nature* **356**, 683–689 (1992).
13. Sadis, S. & Hightower, L. E. *Biochemistry* **31**, 9406–9412 (1992).
14. McCarty, J. S. & Walker, G. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9513–9517 (1991).
15. Brown, C. R., Martin, R. L., Hansen, W. J., Beckmann, R. P. & Welch, W. J. *J. Cell Biol.* **120**, 1101–1112 (1993).
16. Antonino, L. C., Kautz, R. A., Nakano, T., Fox, R. O. & Fink, A. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7715–7718 (1991).
17. Bardwell, J. C. A. & Craig, E. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 848–852 (1984).
18. Schmid, S. L., Braell, W. A. & Rothman, J. E. *J. biol. Chem.* **260**, 10057–10062 (1985).
19. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
20. Liberek, K., Marszałek, J., Ang, D. & Georgopoulos, C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2874–2878 (1991).

ACKNOWLEDGEMENTS. We thank G. C. Walker and J. S. McCarty for DnaKT199A. This work was supported by a grant from the NIH.

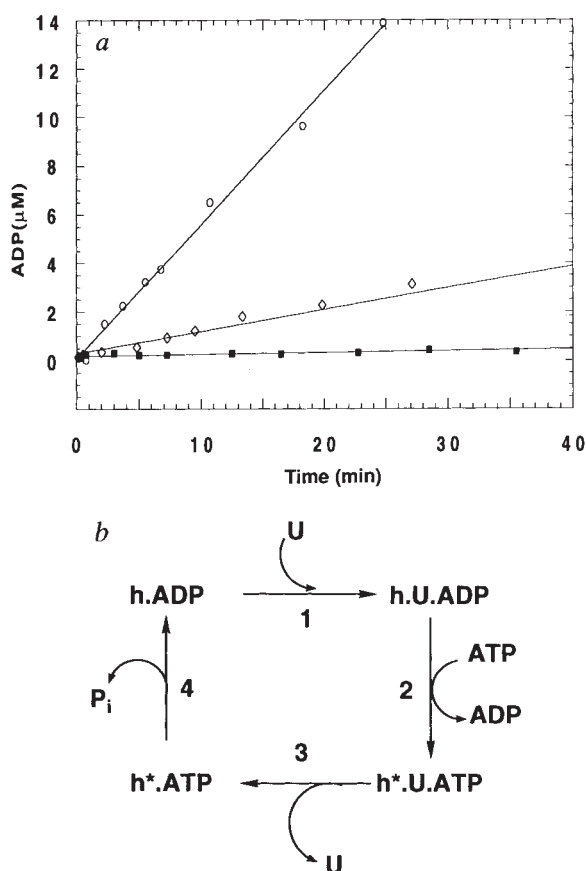


FIG. 3 a, ATPase activity of DnaK at 37 °C. Open symbols, burst experiment; complex between RCMLA and DnaK was formed by mixing aliquots of RCMLA (340 μM in HCl, pH 2), DnaK (3.4 μM in 20 mM Tris-HCl, 10 mM NaCl, pH 6.7), KCl stock solution (2 M in buffer A: 20 mM Tris-HCl, 20 mM NaCl, 10 mM MgCl_2 , pH 7.2) and buffer A and incubating for 10 min at 37 °C; Mg-ATP stock solution (500 μM Na_2ATP in buffer A) was added, then 1 μCi $\text{Na}_2[2,8\text{-}^3\text{H}]\text{ATP}$ (30 Ci mmol^{-1} ; 33 μM in ethanol:water (1:1) from NEN) was incorporated and the ATPase activity followed; final volume was 50 μl . Aliquots were retrieved at the times shown and the ^3H -labelled nucleotides were separated by TLC on cellulose-PEI plates and analysed as described¹¹; final concentrations were: [RCMLA], 41 μM ; [K^+], 200 mM; (○) [DnaK], 2.4 μM ; [ATP], 20.7 μM ; (◇) [DnaK], 0.5 μM ; [ATP], 10.7 μM (identical results were obtained when [ATP] was 21 μM , indicating substrate saturation); (■) DnaK T199A stock solution (1.94 μM in 20 mM Tris-HCl, 20 mM NaCl, pH 7.2); [DnaK T199A], 0.66 μM ; [ATP], 10.7 μM ; rest as before. b, Cycle for the nucleotide-modulated binding and dissociation of unfolded proteins to Hsp70; see text for details.

A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells

Michael I. Bukrinsky^{*†}, Sheryl Haggerty^{*}, Michael P. Dempsey^{*}, Natalia Sharova^{*}, Alexei Adzhubei[‡], Lynn Spitz[§], Paul Lewis^{||}, David Goldfarb[§], Michael Emerman^{||} & Mario Stevenson^{*¶}

^{*} Department of Pathology and Microbiology and

[¶] Eppley Institute for Research, University of Nebraska Medical Center, Omaha, Nebraska 68198–5120, USA

[‡] Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, London WC2A 3PX, UK

[§] Department of Biology, University of Rochester, Rochester, New York 14627, USA

^{||} Program in Molecular Medicine, Fred Hutchinson Cancer Center, Seattle, Washington 98104, USA

PERMISSIVENESS of the host cell to productive infection by onco-retroviruses is cell-cycle dependent¹, and nuclear localization of viral nucleoprotein preintegration complexes will occur only after cells have passed through mitosis². In contrast, establishment of an integrated provirus after infection by the lentivirus HIV-1 is independent of host cell proliferation^{3–5}. The ability of HIV-1 to

[†] Present address: Picower Institute for Medical Research, 350 Community Drive, Manhasset, New York 11030, USA.

[#] To whom correspondence should be addressed.

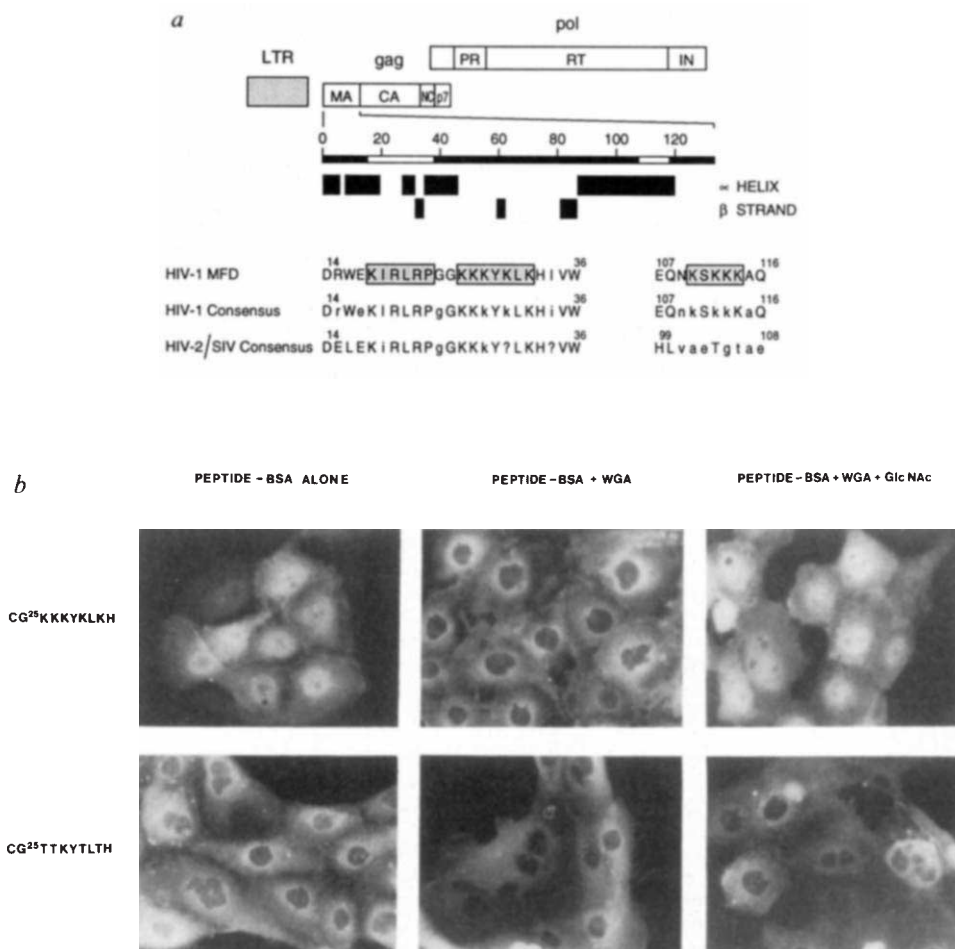
replicate in non-dividing cells is partly accounted for by the karyophilic properties of the viral preintegration complex which, after virus infection, is actively transported to the host cell nucleus. Here we report that the *gag* matrix protein of HIV-1 contains a nuclear localization sequence which, when conjugated to a heterologous protein, directs its nuclear import. In addition, HIV-1 mutants containing amino-acid substitutions in this nuclear localization signal integrate and replicate within dividing but not growth-arrested cells, and thus display a phenotype more representative of an onco-retrovirus.

Nuclear localization signals (NLSs) have been identified for many karyophilic proteins of viral and cellular origin and generally resemble either the single basic domain SV40 large T-antigen NLS (PKKKRKV) or the double basic domain nucleoplasmin NLS (KRPAATKKAGQAKKKK, single-letter amino-acid code)⁶. Sequences within HIV-1 matrix protein with homology to the single domain motif are shown in Fig. 1a. Synthetic peptides containing putative NLSs of HIV-1 matrix were fluoresceinated and conjugated to bovine serum albumin (BSA)⁷. When microinjected into the cytoplasm of PtK-1 cells⁸, a peptide encompassing amino acids 25–33 of HIV-1 matrix directed the partial localization of BSA to cell nuclei (Fig. 1b). This nuclear import was abolished by the co-injection of wheat-germ agglutinin (WGA), an inhibitor of nuclear pore-complex-mediated import, but proceeded in the presence of WGA and its ligand, *N*-acetylglucosamine (GlcNAc). No nuclear localization was observed when BSA was conjugated to peptides containing four lysine to threonine substitutions (Fig. 1b) or to peptides based on a second putative NLS (amino acids 107–116 of matrix) enriched in basic amino acids (not shown).

The *in situ* microinjection approach does not take into consideration the stoichiometry of viral preintegration complex components. As shown for both strong and weak NLSs⁹, increasing

the number of NLS motifs associated with a particle improves the rate and extent of its import. In addition to the valency effects, the size of the carrier molecule greatly influences the extent to which nuclear import of NLS peptide-carrier conjugates are imported to the nucleus. Thus, sequences sufficient to direct nuclear import of BSA (diameter 70 Å) may be incapable of directing large particles such as a retroviral preintegration complex (diameter ~300 Å) to the nucleus. Therefore, the effects of amino-acid substitutions in the NLS of HIV-1 matrix were examined in the context of the viral preintegration complex. Isogenic HIV-1 variants containing amino-acid substitutions within the N-terminal NLS (residues 25–33) of HIV-1 matrix were generated by site-directed mutagenesis. A single Lys 27→Thr substitution and double Lys 26, 27→Thr substitutions within HIV-1 matrix were created and reintroduced into an infectious HIV-1 molecular clone, HIV-1 MFD⁹. We had previously speculated that the ability of HIV-1 to replicate in non-proliferating cells might be due to nuclear entry of the preintegration complex before mitosis^{4,5}. Therefore, the replication of HIV-1 matrix mutants in proliferating compared with non-proliferating cells was evaluated in an integration-dependent transactivation assay¹⁰. The target cells, HeLa CD4-LTR- β -gal cells, were arrested at the G2 phase of the cell cycle by γ -ray irradiation, and the infection of non-proliferating cells relative to infection of untreated cells was measured in a single round of infection⁵. As previously shown⁵, the onco-retrovirus MuLV very poorly infects non-proliferating cells compared with HIV-1 (Fig. 2). In contrast to wild-type HIV, neither the single Lys 27→T substitution (M16), nor the double Lys 26, 27→Thr substitution (M17) was able efficiently to infect G2-arrested cells (Fig. 2). This pattern of virus replication in dividing compared with non-proliferating host cells was more representative of an onco-retrovirus phenotype.

FIG. 1 Putative nuclear targeting sequences within the HIV-1 *gag* matrix protein. a, Boundaries of the major products of the *gag* and *pol* polyproteins are indicated. The positions of predicted α -helices and β -strands were determined as in ref. 21. Sequences that resemble common NLS motifs⁶ are boxed and stippled. Amino-acid numbering, alignment and description of consensus sequences was helped by information contained within the HIV-1 sequence data base²². Upper case letters represent invariant residues whereas conserved and non-conserved amino acids are denoted by lower case letters and the question mark, respectively. b, Subcellular location of HIV-1 matrix NLS peptide-BSA conjugates after microinjection into rat epithelial cells. METHODS. PtK1 cells (rat kangaroo kidney epithelia) were grown on glass cover slips in F-12 medium supplemented with 10% fetal bovine serum. Cells were microinjected in Hank's balanced salt solution, essentially as detailed elsewhere⁸. Fluoresceinated HIV-1 matrix peptide-BSA conjugates containing an average of 6 peptides per BSA molecule⁷ were injected with or without WGA alone (2.5 mg ml⁻¹) or WGA and its ligand, GlcNAc (0.5 M). Microinjected cells were incubated for 30 min at 37 °C, fixed and photographed⁸.



A comparison of the abilities of wild-type HIV-1 and isogenic HIV-1 matrix mutants to elicit a spreading infection in proliferating CD4⁺ cells and in CD4⁺ cells arrested in the cell cycle by aphidicolin is shown in Fig. 3. In proliferating CD4⁺ cultures (–Aph), replication kinetics of both single and double HIV-1 matrix mutants was indistinguishable from that of wild-type virus. In contrast, under conditions of growth arrest (+Aph), virus antigen production in cultures infected with both HIV-1 NLS matrix mutants was not detected (Fig. 3). When the aphidicolin block was removed and cell proliferation resumed, virus replication and antigen production were detected in cultures infected with both HIV-1 matrix mutants (not shown).

Major complementary DNA products of HIV-1 reverse transcription in the cytoplasm and nucleus of acutely infected cells are illustrated in Fig. 4a. Viral cDNA synthesis in dividing compared with growth-arrested cells infected with wild-type HIV-1 and an HIV-1 matrix mutant (Lys 26, 27→Thr) was equivalent (Fig. 4b). Nuclear localization of viral DNA, shown by abundance of nucleus-specific two long terminal repeat (LTR) circle forms of viral DNA, was equivalent in dividing and growth-arrested cells infected with wild-type HIV-1. In contrast,

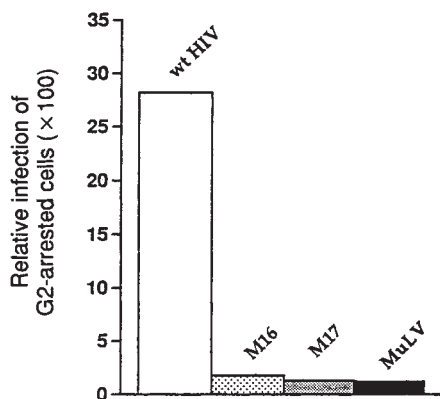


FIG. 2 Relative infection of G2-arrested HeLa CD4 cells by HIV-1, MuLV and HIV-1 NLS matrix mutants. HeLaCD4-LTR/ β gal cells were arrested in G2 by γ -irradiation, and infection was assayed by transactivation of an endogenous LTR- β -galactosidase gene as described⁵. In this assay, HeLa cells, which have been engineered to express high levels of the HIV receptor, CD4, contain a single copy of the bacterial β -galactosidase gene under transcriptional control of the HIV-1 LTR. On HIV-1 infection and integration β -galactosidase is produced after transactivation of the LTR by *tat* protein. Infected cells are scored after visualization of β -galactosidase activity *in situ* by staining cells with x-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside). The relative infection of G2-arrested cells is the number of infected cells in the γ -irradiated (non-proliferating) culture divided by the number of infected cells in the proliferating culture for each viral infection. wt HIV is the parental MFD strain⁹; M16 is the single Lys 27→Thr matrix mutant of HIV-1 MFD; M17 is the double Lys 26, 27→Thr mutant of HIV-1 MFD; and MuLV is a MuLV-based retrovirus vector that encodes the HIV-1 *tat* gene⁵. METHODS. Amino-acid substitutions within HIV-1 *gag* matrix were introduced by site-directed mutagenesis. A 3-kilobase (kb) *Sph*I fragment of the infectious molecular clone, HIV-1 MFD⁹ (similar in genotype to the prototypic HXB2 isolate), was inserted in M13mp19. Single (Lys 27→Thr) and double (Lys 26, 27→Thr) amino-acid substitutions were created by annealing M13 phage DNA with the mutagenic oligonucleotides, M16(5'-aggcctgggggaagacaaaatat-3') and M17(5'-aggcctgggggaagacaaaatat-3'), respectively. Both oligonucleotides were designed to introduce translationally silent *Stu*I recognition sites. DNA fragments containing the desired mutations were inserted in the HIV-1 MFD DNA backbone. Wild-type and mutant HIV-1 variants were recovered 48 h after transfection of HeLa cells and quantified by HIV-1 *gag* p24 ELISA. When cells are transfected with infectious DNA clones, events preceding provirus establishment are bypassed and only late events including assembly and release of virion particles are reconstituted. The amino-acid substitutions within the HIV-1 *gag* matrix NLS did not influence virus maturation and both quantity and morphology of virus particles produced from HeLa cells transfected with matrix mutant and wild-type molecular clones of HIV-1 were identical (not shown).

after infection with the matrix mutant, formation of two LTR circle forms was evident in dividing cells but not growth-arrested cells. Synthesis and nuclear import of viral DNA in proliferating and growth-arrested cells infected with a single (Lys 27→Thr) HIV-1 matrix mutant was identical to that for the double matrix mutant (not shown). Taken together, these results suggest that under conditions of growth arrest, amino-acid substitutions within HIV-1 matrix antigen interrupt the ability of viral nucleic acids to enter the nucleus and as a consequence restricts the ability of HIV-1 to elicit a spreading infection in non-proliferating host cells.

The study outlined here is pertinent to the understanding of determinants which govern the ability of HIV-1 to replicate within non-proliferating cells. This pattern of HIV-1 replication is supported by *in vitro* studies demonstrating provirus establishment in G2-arrested CD4⁺ HeLa cells⁵ and primary macrophages^{3,11} and is reflected by distribution of the virus in non-T-cell compartments of the infected host, such as terminally differentiated and non-proliferating cells of macrophage or microglial phenotype^{12–15}. The propensity for replication in non-dividing cells distinguishes HIV-1, and perhaps lentiviruses in general¹⁶, from the animal onco-retroviruses^{1,2}. Although retroviruses such as MuLV will infect non-proliferating cells and initiate viral DNA synthesis¹⁷ within the context of a preintegration complex, the large physical size of retroviral preintegration complexes¹⁸ precludes them from passing through the aqueous channel of nuclear pores, requiring instead that they enter the nuclear compartment after breakdown of the nuclear membrane during mitosis². The requirement for mitosis at cell division is obviated by the karyophilic properties of the preintegration complex of HIV-1, which undergoes active nuclear import after acute virus infection of non-proliferating cells⁴. Determinants which govern nuclear import characteristics of HIV-1 DNA are contained within the virus core because MuLV pseudotypes containing *gag/pol* proteins of HIV-1 will integrate within non-proliferating cells⁵. Of the *gag/pol* proteins, integrase^{19,20} and matrix²⁰ proteins of *gag* and reverse transcriptase protein of *pol*²⁰ have been identified in association with viral nucleic acids in the context of the viral preintegration complex. We have shown here that the HIV-1 *gag* matrix protein, by virtue of a nuclear localization sequence at its N terminus, contributes to the karyophilic properties of the viral preintegration complex and influences the ability of the virus to replicate within non-proliferating cells. □

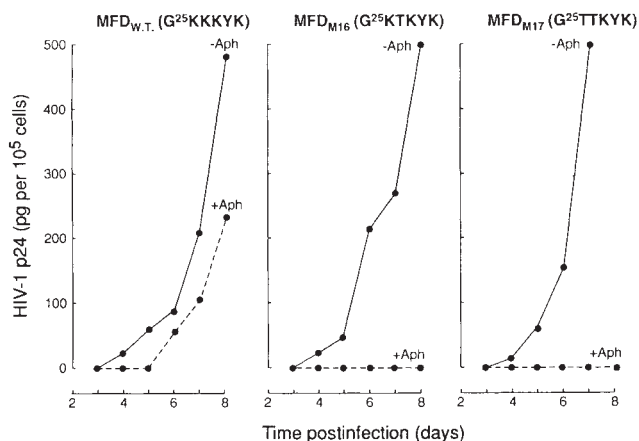
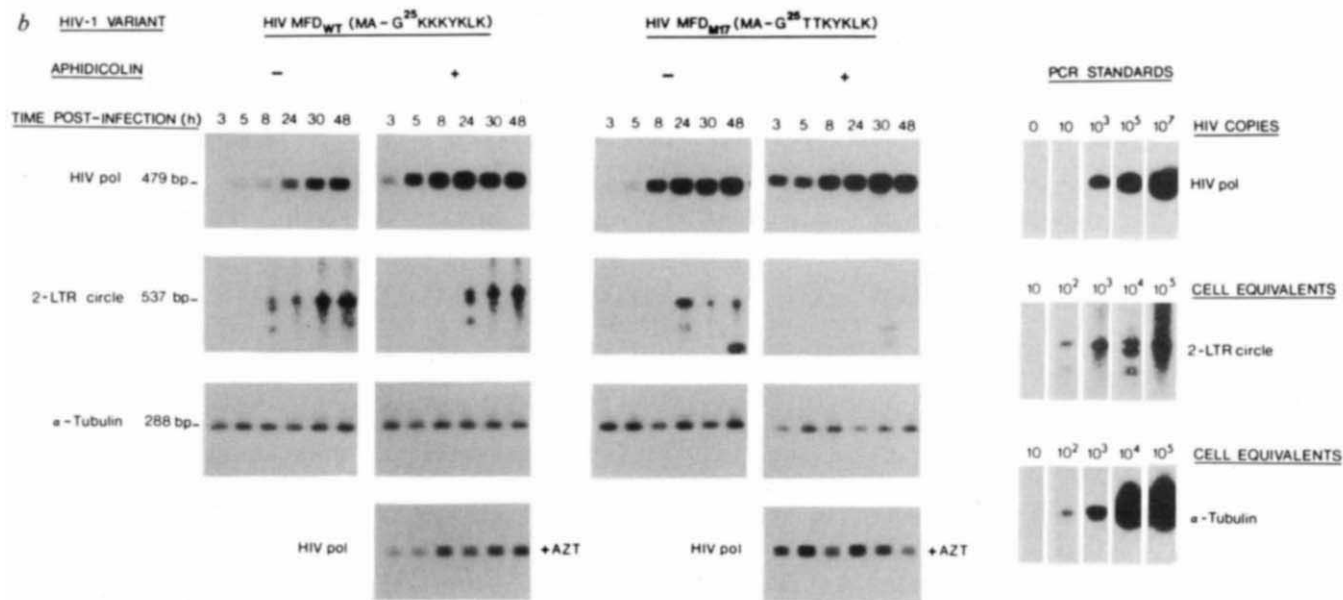
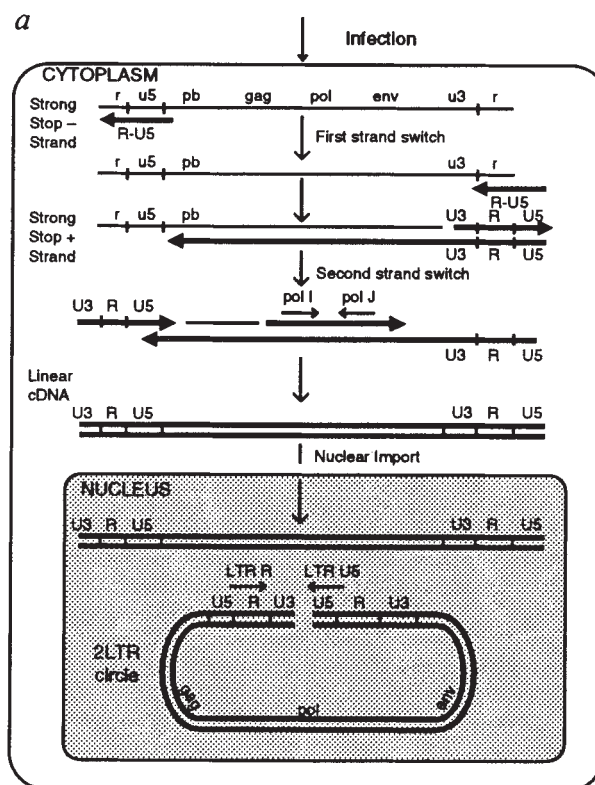


FIG. 3 Replication of wild-type and *gag* matrix mutants of HIV-1 in dividing and in growth-arrested CD4⁺ cells. CD4⁺ MT-4 cell cultures were arrested at the G1/S interface of the cell cycle by the action of aphidicolin as detailed elsewhere⁴. Aphidicolin was removed 72 h after infection with HIV-1 (50 pg per 10⁶ cells). At 24-h intervals, 1 × 10⁵ cells were removed for quantification of cell-associated HIV-1 *gag* p24 in proliferating (solid line) and in growth-arrested (hatched line) cultures.

FIG. 4 Synthesis and nuclear import of viral DNA by wild-type and *gag* matrix mutants of HIV-1 in proliferating and growth-arrested CD4⁺ T-cell cultures. **a**, Major reverse transcription products in cytoplasm and nucleus of HIV-1-infected cells. After reverse transcription is initiated, two strand-switching events result in a full-length cDNA containing duplicated U3, R and U5 regions. After transport to the nucleus, some linear molecules undergo circularization, yielding 2 LTR circle forms. Although such circle forms of viral DNA do not appear to be precursors of the provirus²³, they are formed only after synthesis of full-length viral cDNA and its transport to the nucleus^{1,23} and thus represent a specific marker with which to follow nuclear localization of viral DNA. HIV-1 *pol* I and J primers amplify products of reverse transcription which follow the second strand switch²⁴, whereas LTR R and U5 primers amplify 2 LTR circle forms of extrachromosomal viral DNA. Viral RNA and cDNA are represented by thin lines/lower-case letters and thick lines/upper-case letters, respectively. pb, Primer binding site. **b**, Proliferating (– Aph) and growth-arrested (+ Aph) CD4⁺ MT-4 cells were infected with HIV-1 as outlined in Fig. 3. At the indicated times postinfection, cell aliquots were withdrawn (2×10^5 cells) for isolation of total cellular DNA as detailed elsewhere²⁵. HIV-1 *pol*-specific amplification products obtained in the presence of AZT (1 μ M) represent incomplete reverse transcription products contained within the virion and not *de novo* synthesis of viral DNA^{26,27}.

METHODS. PCR conditions for amplification of linear and circle forms of viral DNA were exactly as detailed elsewhere^{4,25}. One and two rounds of PCR were used for *pol* and LTR circle primers, respectively. PCR amplification products were visualized after Southern blot transfer and hybridization with [α -³²P]-labelled oligonucleotide probes. HIV-1 *pol* standards were generated by PCR on log dilutions of cloned HIV-1 MF DNA whereas standards for circle viral DNA forms and cellular DNA generated by PCR on log dilutions of HIV-1-infected MT-4 cells.



Received 30 March; accepted 2 August 1993.

- Varmus, H. E. & Swanstrom, R. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N., Varmus, H. E. & Coffin, J.) 369–512 (Cold Spring Harbor Laboratory Press, New York, 1984).
- Roe, T., Reynolds, T. C., Yu, G. & Brown, P. O. *EMBO J.* **12**, 2099–2108 (1993).
- Weinberg, J. B., Mathews, T. J., Cullen, B. R. & Malim, M. H. *J. exp. Med.* **174**, 1477–1482 (1991).
- Bukrinsky, M. I. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6580–6584 (1992).
- Lewis, P., Hensel, M. & Emerman, M. *EMBO J.* **11**, 3053–3058 (1992).
- Forbes, D. J. *A. Rev. Cell Biol.* **8**, 495–529 (1992).
- Goldfarb, D. S., Gariepi, J., Schoolnik, G. & Kornberg, R. D. *Nature* **322**, 641–644 (1986).
- Breeuwer, M. & Goldfarb, D. S. *Cell* **60**, 999–1008 (1990).
- Stevenson, M. et al. *J. Virol.* **64**, 3792–3803 (1990).
- Kimpton, J. & Emerman, M. *J. Virol.* **66**, 2232–2239 (1992).
- Gendelman, H. E. et al. *J. Virol.* **58**, 67–74 (1986).
- Shaw, G. M. et al. *Science* **226**, 1165–1170 (1984).
- Koenig, S. et al. *Science* **223**, 1089–1093 (1986).
- Wiley, C. A., Schrier, R. D., Nelson, J. A., Lampert, P. W. & Oldstone, M. B. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7089–7093 (1986).
- Price, R. W. et al. *Science* **239**, 586–592 (1988).

- Haase, A. T. et al. *Virology* **119**, 399–410 (1982).
- Miller, D. G., Adam, M. A. & Miller, A. D. *Molec. cell. Biol.* **10**, 4239–4242 (1990).
- Bowerman, B., Brown, P. O., Bishop, J. M. & Varmus, H. E. *Genes Dev.* **3**, 469–478 (1989).
- Farnet, C. M. & Haseltine, W. A. *J. Virol.* **65**, 1990–1915 (1991).
- Bukrinsky, M. I. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6125–6129 (1993).
- Eliopoulos, E. E. et al. *Int. J. biol. Macromol.* **4**, 263–275 (1982).
- Meyers, G. et al. *Human Retroviruses and AIDS* (Los Alamos National Laboratory, New Mexico, 1992).
- Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. *Cell* **49**, 347–356 (1987).
- Hsu, W.-S. & Temin, H. M. *Science* **250**, 1227–1232 (1990).
- Bukrinsky, M. I., Stanwick, T. L., Dempsey, M. P. & Stevenson, M. *Science* **254**, 423–427 (1991).
- Lori, F. et al. *J. Virol.* **66**, 5067–5074 (1992).
- Trono, D. *J. Virol.* **66**, 4893–4900 (1992).

ACKNOWLEDGEMENTS: We thank T. Pushkarayya for technical assistance, H. Varmus and S. Rhode for discussion and C. Bowers for manuscript preparation. S.H. is a scholar of the American foundation for AIDS research, AmFAR. This work was supported by the NIH (M.S., D.G., M.E.) and by AmFAR (M.B.).

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature*'s staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.

Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing

B. A. Larder, A. Kohli, P. Kellam, S. D. Kemp, M. Kronick and R. D. Henfrey

A comparison has been made between manual and automated DNA sequencing procedures to evaluate the ability to distinguish mixtures of wild-type and mutant sequences. Quantitative detection of such mixtures of HIV-1 drug resistance mutations was best achieved using an automated system that uses fluorescent-labelled sequencing primers. This procedure has a wide range of applications in clinical research, including heterozygote analysis. Software that automatically reports mixed-base positions is presented.

THE availability of rapid, high-throughput automated DNA sequencing technology¹ has obvious applications in clinical research, including the detection of, variations in virus populations², and mutations responsible for drug resistance in virus genomes. However, analysis of clinical samples by manual sequencing or polymerase chain reaction-(PCR) based point mutation assays has revealed that complex mixtures of wild-type and mutant HIV-1 genomes can occur during drug therapy³. Therefore, to assess the likely susceptibility of a virus population to a particular drug during therapy, it would be desirable to perform DNA sequence analysis that can simultaneously quantitate several resistance mutations in multiple genomes. We have carried out comparisons of manual and automated DNA sequencing to investigate the limits of resolution of mutation detection at specific codons in predetermined mixtures of HIV-1 sequences.

In order to model the mixed populations of HIV-1 encountered in clinical samples (for example, DNA extracted from periph-

eral blood mononuclear cells) we firstly made mixtures of HIV-1 reverse transcriptase (RT) DNA cloned into M13 bacteriophage. We mixed defined amounts of each clone, which contained wild-type or mutant residues in RT at codons 181 (TAT = wild-type: Tyr, TGT = mutant: Cys), 215 (ACC = wild-type: Thr, TTC = mutant: Phe) and 219 (AAA = wild-type: Lys, CAA = mutant: Gln). The mutations at codons 215 and 219 confer resistance to AZT^{3,4} and the mutation at codon 181 causes broad resistance to non-nucleoside RT inhibitors such as nevirapine⁵. Mixtures consisted of: 100% wild-type, 90%:10% wild-type to mutant, 75%:25%, 50%:50%, 25%:75%, 10%:90% and 100% mutant. Each DNA sample (4.5 ng total) was then subjected to PCR amplification to obtain sufficient amounts of DNA fragment that encompass the RT codons under investigation.

Manual DNA sequencing

The PCR products of 731 base pairs (bp) were obtained by amplification with oligonucleotide primers comb 2 and comb 3 (see Fig. 1 legend for sequences). After purification using GeneClean (BIO 101), the products were alkali-denatured and then sequenced manually using an internal oligonucleotide primer B (see Fig. 2 legend for sequence) and T7 DNA polymerase (Sequenase, USB) according to an established procedure⁶. The results of this analysis are shown in Fig. 1 for the region of HIV-1 that includes codon 181. It is clear from the autoradiograph that mixed bases at this codon can be seen in the approximate proportions expected from the origi-

nal DNA mixtures. The limit of heterozygote detection under these conditions was a mixture of 75%:25%. For codons 215 and 219 mixtures could be discerned but quantitation was difficult due to the close proximity of bands at this point in the autoradiogram.

Automated DNA sequencing

Two different procedures were used for automated sequencing of the regions of interest in the DNA mixtures. Firstly, an identical 731-bp PCR product was amplified, with the exception that the 3'-oligonucleotide primer comb 3 was biotinylated to facilitate purification of single-stranded (negative sense) DNA using a magnetic bead separation system (Dynabeads M-280 streptavidin, Dynal). Sequencing reactions were performed using the PRISM Sequenase fluorescent dye-labelled dideoxynucleotide system⁷ (Applied Biosystems). Secondly, a smaller PCR product was produced using biotinylated 3'-primer comb 3 with a 5' primer based on primer B that had been tailed with a 18-base universal primer sequence (see Fig. 2 legend for sequence). After magnetic bead separations of single-stranded DNA product, sequencing reactions were performed using both the PRISM Sequenase dye terminator system (ABI), and the PRISM Sequenase fluorescent dye-labelled primer system (ABI). As the universal primer sequence had been incorporated into the PCR product, this could be sequenced with dye-labelled universal primer. Sequence analysis of dye-labelled products was performed using the Applied Biosystems Model 373A DNA sequencer.

Individual lane files of interest were transferred to the Factura software analysis package (ABI). This new package processes files in a batch mode to identify likely heterozygous positions. The positions are marked as potentially heterozygous or of mixed-base composition if the second most intense peak at a particular base position is more than a set percentage (as defined by the user) of the most intense peak. The files were then transferred to the Sequence Navigator software package (ABI), where the individual

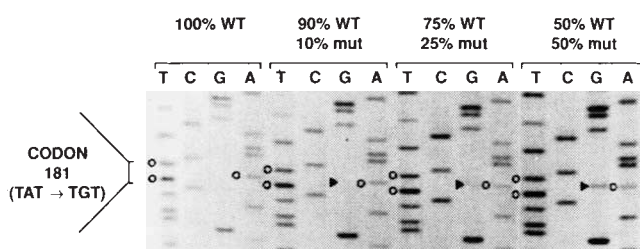


FIG. 1 Autoradiograph showing manual DNA sequence analysis of wild-type and mutant mixture of codon 181 in reverse transcriptase (RT). Mixtures of DNA clones containing wild-type (TAT) or mutant (TGT) sequence at RT codon 181 were made in the proportions described in the text. A total of 4.5 ng of each mixture was subjected to PCR amplification using the 5'-oligonucleotide primer comb 2 (5'-CTGTACCAAGTAAATTAAGCCAGG) and 3'-oligonucleotide primer comb 3 (5'-ATAGGCTGTACTGTCCATTTATCAGG) and PCR conditions as described previously³ for primer pair A/NE-1. PCR products were purified using a glass-milk procedure (GeneClean), alkali-denatured (0.2 M NaOH plus 0.2 mM EDTA, 30 min at 37 °C) neutralized (0.3 M sodium acetate, pH 4.8) and sequenced using T7 DNA polymerase (Sequenase) as described previously⁶. The samples shown on the autoradiograph are wild-type at codon 181 or with 10%, 25% and 50% mutant sequence, respectively. The wild-type bands (TAT) are marked in each lane (O) as is the position of the mutant G band (▶).

FIG. 2 Chromatograms showing automated DNA sequence analysis of wild-type and mutant mixture of codon 181 in reverse transcriptase (RT). Mixtures of DNA clones containing wild-type (TAT) or mutant (TGT) sequence at RT codon 181 were made in proportions described in the text. *a*, A total of 4.5 ng of each mixture was subjected to PCR amplification using the 5'-universal tailed (underlined) oligonucleotide primer B (5'-TGTAACGACGCGCAGTGGATGGAAGGATCACC) and the 3'-oligonucleotide primer comb 3 biotinylated at the 5' end (see Fig. 1 legend for sequence) and PCR conditions as described in Fig. 1. PCR products were purified using Centricon 100 columns (Amicon), and immobilized onto streptavidin-coated magnetic beads (Dyna) followed by strand separation according to the manufacturer's instructions. Eighty per cent of the immobilized products were taken for the sequencing reaction using the PRISM Sequenase fluorescent dye-labelled primer kit (ABI) according to the manufacturer's protocol. The samples shown are wild-type at codon 181 or with 10%, 25% and 50% mutant sequence, respectively. When the relative intensity of the two most intense peaks at a particular base position is greater than some pre-set (user-defined) threshold, a mixed-base position is identified. This can be seen in the data for the 75% and 50% mixture, where the R (IUB code) indicates a A/G heterozygote. *b*, Each mixture was subjected to PCR amplification using primer sets as described in Fig. 1, except the comb-3 primer was biotinylated. The PCR products were purified and immobilized as described in Fig. 2*a*. Sequencing was performed on 80% of the immobilized products using the PRISM Sequenase fluorescent dye-labelled dideoxynucleotide kit (Applied Biosystems) in conjunction with an unlabelled oligonucleotide primer B (5'-GGATGGAAGGATCACC) according to the manufacturer's instructions.

files can be compared and true mixed-base positions verified. Note that in Fig. 2 four different files are being compared and that the mixed-base position shows an increasing percentage of the mutant base as the percentage mutant sample (at codon 181) in the analysed mixture was increased. The Table shows the manual computation of wild-type signal/wild-type signal + mutant signal for both Sequenase primer data, and Sequenase terminator data at all the heterozygous codon positions.

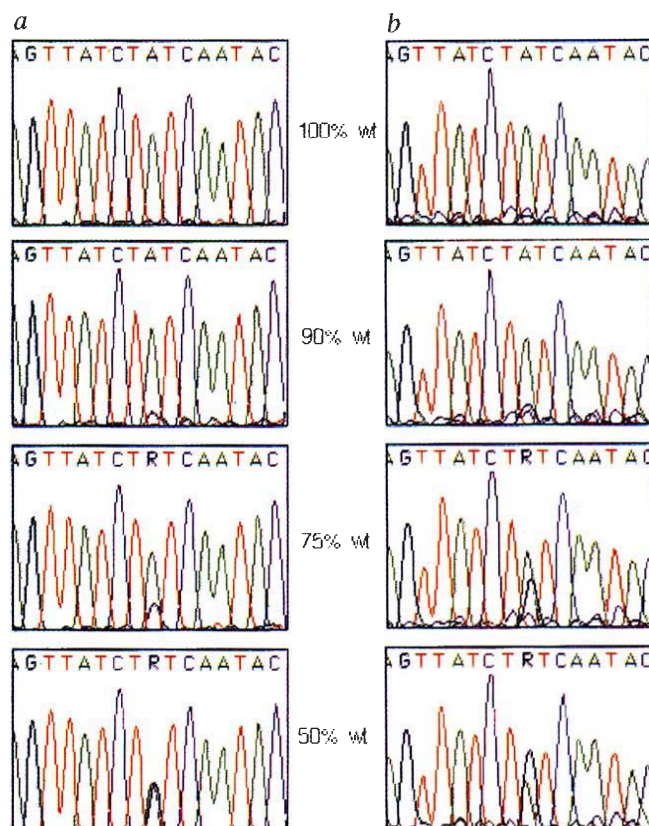
Discussion

The manual and automated Sequenase dye primer sequencing results indicate that the PCR appeared to amplify the mixtures faithfully. The manual sequence data demonstrated that a 25% mixture of wild-type or mutant bases could be detected while 10% mixtures were much less easy to see, whereas the automated Sequenase dye primer data demonstrated that a 25% mutant sample could easily be identified and a 10% level could frequently be detected (see Fig. 2*a* and Table). The benefit that the automated data had over the manual data was the ability of the Factura program to identify automatically the potential heterozygotes, and then the easy visualization of the mixtures using Sequence Navigator. An automated sequencer in conjunction with this software makes it possible to screen a large number of samples for potential heterozygotes, and to semi-automate the positive identification of mixed bases in the sample.

The Sequenase dye terminator data was able to identify a 75%:25% mixture at all the mixed-base positions. However, it was noticed that the 50% mixture did not always produce equal peak intensities for the two

bases, indicating slight differences in the kinetics of incorporation of the labelled dideoxy nucleotide by the enzyme. An interesting observation was that the peak-height distributions for all the fragments created in the reaction showed very consistent data for all the mixed templates (see Fig. 2*b*) regardless of whether the universal primer or primer

B was used (data not shown). This indicates that the dideoxy incorporation at any one position is very consistent and is dependent on sequence motif. Therefore, this chemistry also offers the potential for heterozygote detection, especially when the peak height



Quantitation of peak heights at mixed-base positions using automated Sequenase dye terminator and Sequenase dye primer chemistries

Sequenase primer data

Lane	wt%	c181 A/(A+G)	c215 A/(A+T)	c215 C/(C+T)	c219 A/(A+C)
01	100	0.94	1.00	0.96	1.00
03	90	0.87	0.87	0.87	0.87
05	75	0.74	0.71	0.78	0.65
09	50	0.50	0.48	0.60	0.39
11	25	0.29	0.40	0.31	0.15
13	10	0.11	0.12	0.10	0.08
15	0	0.05	0.00	0.00	0.00

Sequenase terminator data

Lane	wt%	c181 A/(A+G)	c215 A/(A+T)	c215 C/(C+T)	c219 A/(A+C)
03	100	0.92	0.82	0.90	1.00
04	90	0.81	0.80	0.91	0.94
05	75	0.61	0.66	0.72	0.72
06	50	0.38	0.41	0.45	0.32
07	25	0.17	0.18	0.23	0.13
08	10	0.06	0.08	0.12	0.00
09	0	0.00	0.06	0.06	0.00

Peak heights at the critical positions were measured (in arbitrary fluorescent units) by taking a y-coordinate reading with the cross-hair cursor feature in the Sequence Navigator software. The peak-height values were then computed as wild-type signal/wild-type signal + mutant signal.

of each base is known. The major advantages that the Sequenase terminator chemistry offers are the single-tube reaction format, the use of unlabelled primers, and the invisibility of nonspecific terminations.

These data demonstrate the considerable advantage that the automated sequencer has in the identification, quantitation, and visualization of heterozygote mixtures. This technology can be widely applied to the analysis of drug resistance mutations, and any other heterozygote analyses involving DNA sequencing. □

B. A. Larder, A. Kohli, P. Kellam and S. D. Kemp are at the Antiviral Therapeutic Research Unit, The Wellcome Foundation Limited, Langley Court, Beckenham, Kent BR3 3BS, UK. M. Kronick is at Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA and R. D. Henfrey is at Applied Biosystems Limited, Kelvin Close, Birchwood Science Park North, Warrington, Cheshire WA3 7PB, UK. For more information, fill in reader service number 100.

1. Smith, L. M. et al *Nature* **321**, 674-679 (1986).
2. Leitner, T. et al *Biotechniques* **15**, 120-127 (1993).
3. Larder, B. A., Kellam, P. and Kemp, S. D. *AIDS* **5**, 137-144 (1991).
4. Larder, B. A. and Kemp, S. D. *Science* **246**, 1155-1158 (1989).
5. Richman, D. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11241-11245 (1991).
6. Winship, P. R. *Nucleic Acids Res.* **17**, 1266 (1989).
7. Lee, L. G. et al *Nucleic Acids Res.* **20**, 2471-2483 (1992).

ADVERTISEMENTS

Synthetic Oligonucleotides

Rapid Quality Worldwide Service

£2.50 per base, no setup!

option of H.P.L.C. purification
biotinylation, fluorescent labelling
RNA Synthesis

phosphothioates, many other modifications

OSWEL DNA SERVICE

Department of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road,
EDINBURGH EH9 3JJ

Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No.18

**Highest Quality
Oligonucleotides**

Synthesised, purified and shipped in 3 days

£2.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No.56

Biomedical briefs

Product launches for the autumn season include a culture system for human bronchial/tracheal epithelial cells and a new peristaltic pumping system.

CLONETICS introduces BronchialPack, a complete normal **human cell culture system** containing normal human bronchial/tracheal epithelial cells, either cryopreserved or proliferating (*Reader Service No. 101*). Packs also include optimized



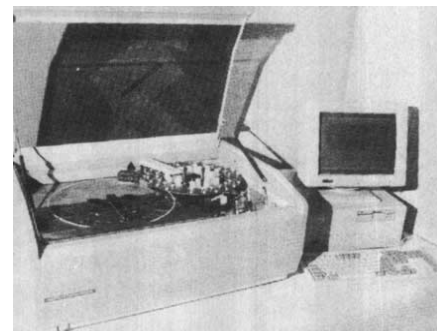
BronchialPack — the complete culture system from Clonetics.

serum-free bronchial epithelial cell growth medium and all the reagents necessary for propagation and subculture *in vitro*. The cultures have applications in respiratory and asthma research, *in vitro* toxicology and basic research.

The new Trizol reagent from Life Technologies is a ready-to-use reagent for the isolation of total RNA, DNA and proteins (*Reader Service No. 102*). The company says that this monophasic solution of phenol and guanidin isothiocyanate offers a number of advantages: a red dye allows for the simple removal of RNA-containing aqueous phase; many samples can be processed simultaneously without compromising RNA quality or quantity; and, typically, RNA can be purified in an hour and DNA or protein in three hours. It also provides a degree of flexibility, and is said to perform equally well with small as well as large quantities of tissue and cells of human, plant or bacterial origin. (1 ml of reagent should be sufficient for isolating total RNA from 100 mg of tissue or 10⁷ cells.) Total RNA isolated by Trizol is free of protein and DNA contamination and can be used for northern blot analysis, dot blot hybridization, poly(A)⁺ selection, *in vitro* translation, RNase protection assays and molecular cloning. Purified DNA and proteins are suitable for Southern blotting and western blotting, respectively. Trizol is available in 100- and 200-ml quantities for US\$95 and US\$175.

Compas 242 is a new type of multiple peptide synthesizer that synthesizes 24

peptides (20 µmol each) simultaneously using Fmoc/tBu chemistry (*Reader Service No. 103*). The instrument, along with the ready-to-use segmental or compartmentalized carriers attached to plastic frames, is now available from Spyder Instruments. The plastic frames are mounted to the perimeter of a centrifugal plate, so that all reagent and wash solutions can be easily removed from the carriers by centrifugation. Four independent gear pumps dispense deprotection and coupling reagents, as well as wash solutions. Amino acid solutions are dispensed by pneumatically operated spray pumps. The company says that the use of solenoid valves and robotic arms for the dispensing of reagent and wash solutions is circumvented, avoiding the potential risks of cross-contamination and mechanical failure. And,



Synthesize 24 peptides in one go with the Compas 242 multiple peptide synthesizer.

as only volumes equal to the wetting or inclusion volumes of the solid support are used, the solvent consumption (4 ml per peptide per cycle) is lower than that of most synthesizers. Compas 242 is controlled by a standalone IBM-compatible computer. Upon entering the amino acid sequences of the peptides to be synthesized, the software provides calculations (weights, volumes) for all the amino acids, reagents and solvents needed. The software also provides two standard synthetic protocols, which can be easily changed for special syntheses.

PRO Scientific manufactures a line of laboratory **homogenizers with a new sealed tube chamber** (*Reader Service No. 104*). The "ST" Series tube assemblies allow for sealed homogenization in standard 5-, 16- and 19-ml capacity tubes, as well as 50-ml conical tubes. PRO Scientific says that its new "ST" Series fits on the PRO 200 handheld or postmounted unit, as well as the PRO 300 benchtop unit.

Two new **microscopes**, the BX50, and lower cost alternative, the BX40, are now

ADVERTISEMENTS

Micromanipulators

Easy to Use Systems for:

- Microdissection
- Probing
- Soldering

SINGER
INSTRUMENTS

"A responsibility to Science"

Tel: (+44) 984 40226 Fax: (+44) 984 41166

Singer Instrument Co. Ltd., Roodwater, Watchet, Somerset, England, TA23 0QL, UK

Reader Service No.50

epitope
custom peptides

Unit 308
Newcastle Technopole
Trafalgar Street
Newcastle upon Tyne
NE1 2LA, UK

Tel. UK 091 201 2179 Fax. 091 201 2191

- Custom peptide synthesis
- Protein Sequencing
- Antibody Production
- Computerised Epitope Prediction

FREE Polo-shirt with orders over £400

dti SMART Award Winner 1993

Reader Service No.26

PEPTIDE ANALYSIS & PURIFICATION

Polymeric reversed phase
HPLC columns

- * Long-term mechanical/chemical stability
- * pH range 1 - 14

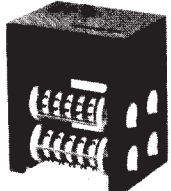
For further details, tel:
USA: (413) 253 9554
UK: 0694 723581

Netherlands: **Polymer Laboratories**
03404-22508

SEPARATION SCIENCE DIVISION

Reader Service No.76

DIABORM®
EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

DIABORM
Gwille
P.O. Box 690126
D-81215 Munich 65, FRG

Reader Service No.45



Olympus's BX50 microscope.

being offered by Olympus (Reader Service No. 105). The BX50 system microscope features a 100 W halogen lamp providing sufficient illumination for most microscopic techniques and photomicrography applications. The BX40 features 30 W halogen illumination, which is sufficient for brightfield, darkfield, phase-contrast and DIC microscopy. The BX50 comes with the new UIS (universal infinity system) infinity-corrected optical system that has an extended field of view (F. N. 22 + F. N. 26.5), in addition to a complete range of accessories that allow it to be adapted for all advanced microscopical applications. Olympus is aiming at the clinical market with the BX40.

Reagents

D609, which is available from Kamiya Bio-medical, is a xanthate compound that is a highly specific **phosphatidylcholine-specific phospholipase C inhibitor** (Reader Service No. 106). The compound also prevents the activation of protein kinase C. As D609 specifically inhibits phosphatidylcholine-specific phospholipase C, a key enzyme in transcription induction of proto-oncogenes such as *c-fos* and of NF- κ B in HIV-1, it should find many applications in signal transduction research. It also has antitumour and antiviral activities and can be used in tissue culture systems.

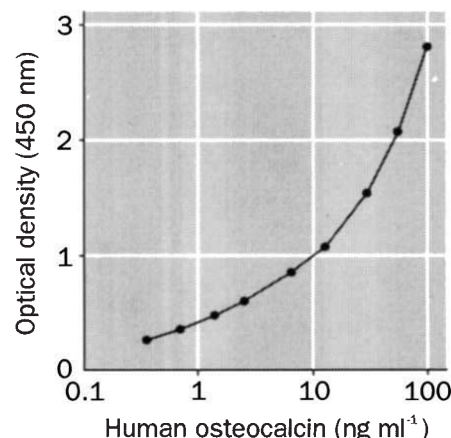
Newly available from Genzyme are **recombinant human MIP-1 α and MIP-1 β** (Reader Service No. 107). Recent studies have shown that both recombinant MIP-1 α and MIP-1 β bind to and induce chemotaxis of monocytes/macrophages and T cells, but not of neutrophils. MIP-1 α may be used to inhibit haematopoietic stem cell growth, whereas MIP-1 β has been shown to counteract this inhibition. With a stated purity of 99 per cent, recombinant human MIP-1 α and MIP-1 β can be used in the research of cell-specific chemotaxis, the inflammatory response and regulation of haematopoietic stem cell proliferation.

Angiotensin II is a peptide hormone involved in the water and sodium homeostasis

of the body. Because of its central role in the control of blood volume, problems with the regulation of the angiotensin system have the potential for serious consequences such as hypertension. Angiotensin type-1 receptors appear to mediate most of the known actions of angiotensin II. Chemicon has introduced a new **rabbit polyclonal antibody to angiotensin II type-1 receptor** that can be used to explore the angiotensin type-1 receptor-G-protein interactions (Reader Service No. 108). It may also be used for immunohistochemistry mapping of angiotensin type-1 receptors in angiotensin target tissues.

Assay systems

Biomedical Technologies has introduced an **enzyme immunoassay for intact human osteocalcin** (Reader Service No. 109). Unlike many polyclonal antibodies that detect both intact and fragmented osteocalcin, the BTI kit uses monoclonal antibodies



Intact human osteocalcin sandwich EIA.

directed towards the amino- and the carboxyl-terminal regions of the protein and measures only intact osteocalcin, which is synthesized *de novo* by the osteoblast. The assay involves a three-hour protocol and requires only a 10- μ l sample. The stated sensitivity is 0.5 ng ml⁻¹. BTI also manufactures reagents for bovine, rat and mouse osteocalcin assays.

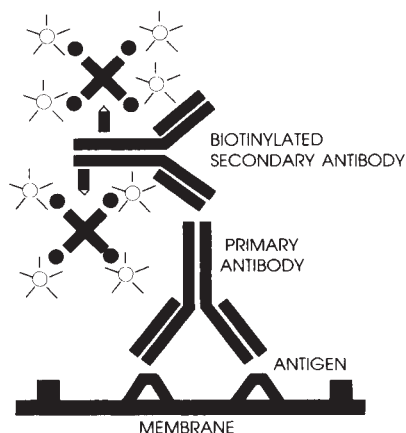
Bender MedSystems has expanded its product range to include six new **ELISA kits** designed for the quantitative detection of the respective soluble molecules in human body fluids (Reader Service No. 110). The line-up includes kits for manganese superoxide dismutase, sTNF-R (80 kDa), sL-selectin (also known as LECAM-1, leukocyte adhesion molecule), s p185^{HER-2} (c-erbB-2), sCD44std and an isoform of CD44var (v6). The assays use precoated microwell strips for convenience and require only room temperature incubation of 2 h 15 min.

Chemiluminescent corner

A new generation of 1,2-dioxetane **enzyme-activated chemiluminescent substrates** for alkaline phosphatase, CDP NATURE · VOL 365 · 14 OCTOBER 1993

and MDP, are now available from Tropix (*Reader Service No. 111*). These two substrates are said to offer a unique combination of properties with respect to overall signal intensity, signal-to-noise ratios, and speed. Tropix says that both MDP and CDP offer good signal-to-noise performance in immunoassay applications with differing characteristics of light emission kinetics. MDP is intended for assays in which speed of determination is of importance; CDP can be used in research applications such as chemiluminescent DNA sequencing. These two new substrates complement the company's CSPD substrate, which is designed for the chemiluminescent detection of nucleic acids and proteins.

UltraLume is the new **immunoblot detection system** from BioGenex Laboratories designed for the chemiluminescent detection of human, mouse or rabbit antibodies in western or dot blots in a procedure that can be completed in under two hours (*Reader Service No. 112*). With a stated sensitivity approaching that of radioactive methods (as little as 10 pg of human IgG), the system



UltraLume — said to provide a sensitivity approaching that of radioactive methods.

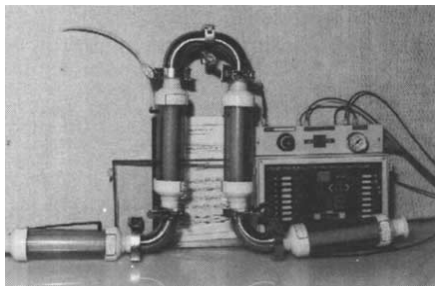
offers the benefit of reducing environmental, health and shelf-life concerns. The system is based on CSPD, a chemiluminescent substrate of alkaline phosphatase. Signal amplification is achieved through the use of a species-specific, biotinylated secondary antibody and alkaline phosphatase-conjugated streptavidin label. The kit can be used to process up to 960 cm² of membrane area.

The Luminograph LB 980 is a high-performance, **low light-level imaging system** from E G & G Berthold that is intended for a wide range of chemiluminescence and bioluminescence applications (*Reader Service No. 113*). These include Southern, northern and western blots, gel electrophoresis, and cell biochemistry studies of bacterial cultures, tissue sections and organisms. The LB 980 system consists of a light-tight sample chamber, a high dynamic range video camera system and a PC-based image

processing system that contains image acquisition and analysis, image enhancement, storage and printing functions.

■ New hardware options

American Dengi has developed a **peristaltic pump** that does not have any rotating or rigid moving components (*Reader Service No. 114*). The device operates on the principle of the sequential collapsing of three or four serially interconnected, flexible tubular pumping segments. Each segment consists of a flexible tubular bladder and a rigid jacket, comprising a "pumping channel" with



American Dengi's pumping system has no rotating or rigid moving parts.

a constant cross-section. The company says that this design virtually eliminates any possible clogging of the pumping mechanism, while the flexible material of the tubing (silicone, for example) tolerates various inserts or solid bodies inside the channel. The pumping process, which is generated by soft flexing of tubular segments, assures reliability and provides pumping pressures that can vary from 3 to 100 p.s.i. Adjustable pressures and a "no-shear" principle of operation allow for either soft, "non-destructive" transfer of "live" organisms or cell cultures, or for the generation of high-pressure sprays or injections. Multiple cartridges with varying internal volumes and made out of different materials can be interconnected in parallel or series, providing flows from millilitres per day up to four gallons per minute. This process can be instantaneously reversed, stopped or reciprocated for mixing, if necessary. Additional coaxial pumping line can be inserted inside the main pumping channel, in which case the latter serves only as an actuator. This technique allows the use of the inserted tubing as a sterile and/or disposable pumping line.

Jouan announces the availability of a new **centrifugal evaporator**, the RC10.09, which is designed to provide users on a budget with an entry-level model (*Reader Service No. 115*). Featuring a transparent Plexiglass lid, Jouan says that the RC10.09 retains many of the advantages of the standard RC10.10 — a large capacity stainless steel bowl, force-field motor with a single moving part and a wide range of rotors for all types of sample container. A version with a PTFE-coated bowl is available for use with strong acids. The RC10.09 is intended for

the concentration of samples before analysis and should be of particular interest to users of HPLC, gas chromatography and electrophoresis for sample preparation.

The Fluostar from SLT Labinstruments is a computer-controlled **fluorometer for fluorescence-based multiwell plate assays** (*Reader Service No. 116*). The instrument captures the emitted light with a side window, current-type photomultiplier and quantifies and characterizes raw data by integration of the current signal during the flash time. All parameters are programmable: excitation and emission wavelengths, temperature, data acquisition modes, result storage options, data processing. This enables users to customize and optimize procedures in applications that include fluorescent immunoassays, DNA and protein estimations, cytotoxicity testing, cell proliferation studies, monoclonal antibody screening and toxicology assays. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

ADVERTISEMENTS

Professional Custom Services

DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

V.G. Services

1200 Wonderland Rd. S. Bldg. #7 Unit #2
London, Ontario Canada N6L 1A6

CALL FAX: (519) 652-3758 ALT. FAX: (519) 652 0053

Reader Service No.32

Teflon[®]
Coated
and

Bare Wire from
MEDWIRE

Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

*Reg. TM of E. I. DuPont & Co

Reader Service No.25

A window of opportunity

Douglas McCormick and John Hodgson

Still in shock from sepsis, biotechnology now has to face healthcare Clinton-style. But research in biotechnology is on the increase as its products surge towards the markets.

AT first glance, things look bad for the biotechnology job market. Tight capital, the virtualization of the American corporation, and jitters over healthcare financing reforms have cast a pall over the sector. In the first half of the year, US biotechnology companies raised just over \$700 million from public offerings, a fraction of what they were bringing in a year or two ago. All but \$60 million of that came in the first quarter: after the dismal performance in Phase III clinical trials of Synergen's lymphokine-receptor antagonist for treatment of septic shock (coming close on the heels of clinical trial debacles for Xoma's and Centocor's drugs for the same indication), the biotechnology capital markets all but dried up.

At the same time, the major pharmaceutical companies have been trimming staff in unhappy anticipation of the Clinton healthcare plan (just now formally released): Johnson & Johnson is eliminating 3,000 jobs; Bristol-Myers Squibb has identified 2,250 positions to cut as it trims its staff by 6%; Syntex plans 1,800 layoffs over the next five years. Marion Merrell Dow is cutting its staff by 1,300; Ciba Geigy's drug division is laying off 550; Cytogen laid off 31% of its staff; Sphinx cut 29%; Rhône-Poulenc Rorer is reshuffling its executive corps, and Baxter has announced a hiring freeze.

And yet.

The good news

And yet research spending continues to rise as biotechnology-based companies, big and small alike, try to fill the product pipeline and ensure smooth passage into the 21st century. The 15 multinational pharmaceutical companies tracked by *Bio/Technology* invested \$11 billion in

product research and development during the last fiscal year. That's 13% more than the year before. The 109 tracked biopharmaceutical companies spent \$1.9 billion on R&D (a 71% increase, following last year's 67% increase)¹. Taken together, the big and small companies have increased their research spending by more than \$2 billion — which translates (using a plausible \$250,000-per-researcher rule-of-thumb) into some 8,000–10,000 primary research jobs and a larger number of technician and support positions.

The reason for these increases is clear: the rising tide of new drugs flowing into the clinic. Back in 1980, the US Food and Drug Administration's Center for Biologics Evaluation and Research (CBER) received fewer than 100 investigational new drug (IND) applications, all of them derived through traditional drug discovery. By 1991, the number of new applications had swelled to about 550, the majority of them products of biotechnology (see figure)².

Today, just a score of biotherapeutics have won approval in the United States, with a few more than that approved in Europe. At last count, though, there were some 50 more in Phase III clinical trials, 100 in Phase II, and 150 in Phase I. Biotechnology companies need new people to bring these new products into the pipeline and to manage clinical trials and manufacturing scale-up. Martin Griffin and Sally Hayward of Project BEMET (Biotechnology in Europe Manpower, Education and Training), in a recent survey of small-to-medium sized biotechnology companies in Europe, showed that over 70% of over 200 companies surveyed had increased their human resources by 25% or more: 40% had increased personnel

by 100%. Moreover, the companies expected that the demand for highly qualified people would continue for the next five years. As one might expect from small research-driven companies, most of them recruited in research and development

Expanding research areas

Steve Aebly is a recruiter, of sorts. His company, Career Connections, in Thousand Oaks, CA, runs job fairs, an increasingly popular mechanism for bringing employers and employees together. More and more of Aebly's business is in biotechnology.

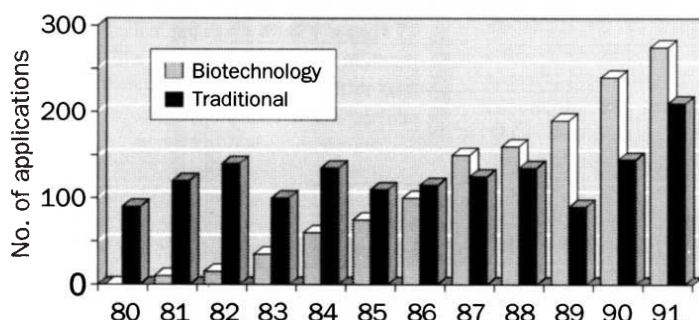
He has talked a lot about industry hiring trends with more traditional recruiters, those who do job training for biotechnology, and especially with personnel managers at biotechnology companies. Demand, he says, is increasing in five areas:

- Small-molecule chemistry
- Clinical trials management
- Quality control/quality assurance
- Plant management
- Process design and scale-up

Small molecules. It sometimes seems that every thoughtful biotherapeutic company has geared up a small but serious small-molecule effort over the past year or so. Synthetic organic chemicals — the backbone of traditional pharmacology — do not suffer proteolysis in the gut; they have a longer shelf-life than proteins; they can penetrate tissues and cells that long-chain peptides cannot; and 25 of the 25 biggest-selling drugs are oral formulations. So, in a sense, biotechnology has come full circle, rejoining the organic chemistry from which first biochemistry and then molecular biology diverged decades ago.

For Xenova, based in Slough in the UK, the discovery and development of small-molecule drugs is its entire business — it has never been involved in protein drugs. The Head of Human Resources there, Alan Garmonsway, says that this year Xenova has increased its head count in all the drug discovery and development areas (microbiology, biochemistry and natural products chemistry), and is now taking on pharmacologists, too. For PhD-level entrants, the company offers both permanent positions and short-term (two-year) contracts. Garmonsway says that short-term contracts can suit both parties: candidates from abroad, for instance, might prefer to make it known at the

Applications to FDA for 'investigational new drug' status



by 100%. Moreover, the companies expected that the demand for highly qualified people would continue for the next five years. As one might expect from small research-driven companies, most of them recruited in research and development

TABLE 1 — JOB CREATION

Industrial focus	% Of new job creation
General pharmaceuticals	22.4%
Human therapeutics	47.2%
Human diagnostics	4.3%
Animal therapeutics	1.6%
Total pharmaceutical	75.5%
Agriculture	5.7%
Bioprocess engineering	3.9%
Instrumentation and supply	2.5%
Informatics systems	2.3%
Textiles	1.6%
Basic research	1.6%
Food and beverage	1.6%
Medical devices	1.4%
Chemicals and enzymes	1.1%
Environment	0.9%
Finance	0.5%
Education	0.5%
Biomaterials	0.5%
Exhibition organizing	0.2%
Energy	0.2%
Law, patent, technology transfer	0.2%
	100.0%

outset that they wished to return home after a fixed period, while the company may need such staff to fulfil a specialist role or to control the growth of its skills base in new areas. Most new appointments at Xenova, however, are to permanent positions.

Towards the market

The rush to market has generated the remaining entries on Steve Aeby's hot-job list. Candidates for these positions tend to hold BSc and MSc degrees, rather than PhDs, and they tend to prefer practical challenges to the lure of the research bench. The jobs generally require juggling technical and regulatory expertise: in industry, satisfying FDA or European regulatory authority demands for proof of safety, efficacy, and process integrity is the name of the game.

Clinical trials. Clinical trials associates collect, coordinate, and safeguard the integrity of clinical trials data. This is no trivial task: poorly designed trials and low-quality data may have caused more heartache in growing biotechnology companies than any other single factor (except, of course, for shortages of cash). Some clinical trials associates work for the

drug manufacturer; others work for clinical research organizations (CROs), specialist bodies which carry out clinical trials under contract to the manufacturer.

Quality control/quality assurance and plant management. As far as the regulatory authorities are concerned, the quality of the product is defined by the process used to make it. Pharmaceutical companies must go to great lengths to keep their processes within the narrow limits prescribed during process validation. Quality control/quality assurance professionals, plant managers, and production technicians are responsible for keeping them that way.

Process design and scale-up. Engineers — typically bachelors and masters level — face the often difficult task of matching the researchers' laboratory system and adapting it to production methods that are at the same time economical, scalable, and acceptable to the regulatory authorities. They know that process steps, once embedded in an IND and sanctified by (expensive) clinical trials, may be difficult or impossible to modify without going back to square one and starting trials all over again.

Among candidates for all of these

positions, Aeby observes, experience is at a premium: "There aren't that many people with ten years of experience in quality control." So biotechnology companies commonly hire candidates with solid academic backgrounds but little experience, and train them to create in-house pools of talent. In keeping with the trend towards virtualization, many companies first bring in new employees as temporaries.

Aeby's impression is that the total number of new jobs in biotechnology remains unchanged, but the mix is clearly changing. "There is less demand for PhDs in protein chemistry," he notes, "and more for BScs and MScs in biology with some experience in some aspect ... any aspect ... of moving products through production and clinical trials." The BEMET data confirm that there is a shift in the pattern of employment, although the rapid overall upward trend in biotechnology employment masks the shift to some extent. The European companies included in the survey anticipated increased demand for immunologists and engineers but a slight slackening in the need for biochemists and microbiologists.

The commercial environment

Regulatory affairs, production and quality control — three of the areas identified by Steve Aeby — are important areas for PhD recruitment at EuroCetus, the Amsterdam-based European subsidiary of Chiron Corporation (Emeryville, California; it acquired nearby Cetus in 1992). Much of EuroCetus' business is in the production, registration and testing of protein drugs developed by the parent company, rather than in developing new therapeutic concepts.

Jaap Bosveld, the Head of Human Resources at EuroCetus, says that recent PhDs will need to adapt to the different demands of commercial research. "People from a university research environment will find that it is a lot different in commercial surroundings. One of the key differences is the need to work under GLP (Good Laboratory Practice) conditions." Another difference is in the imposition of strict objectives and tight timescales. "The deadlines are not set by the research managers within the company," Bosveld explains, "They also come from outside the company — from collaborators or the regulatory authorities — which may require answers within weeks rather than months."

There are also clear differences between the research environment in biotechnology companies and that in larger, 'integrated' companies. Well developed communication skills and flexibility, for instance, are prerequisites for developing teamwork according to Xenova's Garmonsway. "Operating a flat organization and a matrix approach, people who can

Retraining programmes

Jill Hansen is manager of the biomedical manufacturing training programme at the University of California at San Diego extension service. California devotes 0.1% of unemployment tax receipts to retraining programmes. Her programme has obtained a modest grant, about \$300,000, to retrain 90 workers displaced from defence industries for jobs in biotechnology companies. The seven-week course is the product of consultation with participating San Diego

area companies (about a score of them at this point). Here again, the companies need engineering, quality control quality assurance, process validation, and plant operation personnel — along with welding inspectors and technical writers.

There are a number of training programmes scattered about the United States. Useful contacts are state biotechnology associations, state university extension services, or state employment offices.

So you think you want to work in a biotechnology company?

There is no shortage of applicants for new R&D posts at Xenova. "We might get 80 applicants for, say, a natural products chemist position," says the company's Head of Human Resources, Alan Garmonsway, "but of those we will only find around five with the right sort of approach." So how does Xenova narrow down its choice?

One major filter, of course, is the selection of candidates for interview. This usually removes most of those who have not made a choice between the different attractions of academic or industrial research, and those whose scientific interests cannot be aligned with those of the company. The interview itself then refines that subset.

The first part of any interviews at Xenova will probably be with the scientist leading the research team. There will clearly be a scientific discussion but, according to Garmonsway, attributes peripheral to the scientific skills are just as important. "In a few cases, we will look for someone with the specific skills we need for a particular project, especially for long-term projects. But we also need flexibility. For instance, we might hire a cardiovascular pharmacologist but later on we might need them to apply their expertise in oncology, too. We are forever having to tread the line between the specialist and the generalist."

Garmonsway is looking for what he calls the "classic project-management qualities — organisation, teamwork, junior management skills and self-motivation. No doubt most PhD candidates have also been exposed to crisis management, but we do want to find out how well they have planned other aspects." To this end, he will ask how candidates have approached their PhD —

how did they gather the resources and the information they needed? What he would be looking for would be answers which indicated that the candidates had developed the ability to manage their work during their PhD, and that they could operate without a high degree of supervision. There are, after all, some PhD supervisors whose dictatorial style can stifle a graduate student's opportunity to develop independence in research.

Garmonsway also asks how candidates rate themselves as scientists now as compared to two years ago. The answer that interests him the most is along the lines of 'The science is just as exciting, but now I'm getting more out of it because I'm more organised now than I was.'

Interviewees would also spend some time with the department head and, just as important, with some of the other PhD- and graduate-level employees. There would also be a tour of the facilities. "After all," says Garmonsway, "it is as important for the candidate to select us as it is for us to select the candidate."

For the successful candidate, the learning process continues throughout their time with the company. "One of the first things we will do with a new research employee," says Alan Garmonsway, "is to train them in presentation skills. Within 6 months of arriving at Xenova, a new PhD will be involved in project presentation to clients and in internal presentation."

So, if you are a team player, flexible, a good communicator, organised, self-motivated and ... oh, yes ... a good scientist, there may be a place for you in a biotechnology company.

cross boundaries are particularly important." The reward for PhDs coming into a smaller company is one of greater involvement. "At recruitment, things may not seem a lot different," says Garmonsway, "but down the line, the job at Xenova will be much broader than at many larger companies. People will see a wider range of activities. In a larger company, the PhD entrant is at least two levels of organization away from a broader scientific or managerial position." That is part of the attraction of biotechnology: "not many of our people are attracted to compartmentalized jobs in big companies," he says.

Steve Hobbs handles employee recruiting and hiring for Biogen, unique among first-generation biotech companies for

being free-standing and profitable a decade later. His company may be in a stronger position than most, but its employment needs are typical. Right now, Hobbs says, Biogen is hiring new people across the board. Biogen is increasing its research staff by 25% this year, and is looking for a similar increase next year. The new researchers — from senior scientists to recent post-docs — are drawn equally from industry and academic institutions.

Downstream, the company's employment wish-list looks a lot like Aebi's: quality control/quality assurance, process development, clinical trials associates, process operators and manufacturing technicians. Hobbs says that Biogen's Boston/Cambridge-area competitors all seem to be looking for the same kinds of workers — but there are not that many with the desired 3–5 years of experience. Among plant personnel, experience in operating fermentors and doing cell culture is most in demand. The company hires only experienced clinical trials associates; others in the field say trainees with backgrounds in nursing and medical administration may fit the bill.

At the more legal end of the techno-

legal spectrum, Biogen is also looking for regulatory affairs professionals and patent attorneys.

New positions

Each year, *Bio/Technology* magazine surveys its readers to track compensation and employment trends. This year's survey was completed in August and published in the magazine's September issue³. When asked whether they plan to hire new staff in the coming year, 44% of this year's 208 respondents answered in the affirmative. They plan to add 463 new positions to the 2,481 the respondents currently supervise — a 19% increase.

There were few surprises in the pattern of new-job creation (Table 1). Jobs follow money, and the bulk of current investment in molecular biology continues to flow to pharmaceuticals, especially human therapeutics. Agricultural research runs a distant second. The survey also shows relatively rapid growth in bioprocess design and facilities engineering — at least compared to previous levels. And the responses showed a strong increase in interest in applying computational methods to biology. Informatics was expected to contribute about 2.3% of the new jobs in 1994.

What about those who have just entered biotechnology industries? About a quarter of the survey's respondents had been employed for five years or less. Twenty-four percent of these newer employees hold bachelors' degrees. Fifty-two percent hold PhDs, and 9% hold masters. PhDs earn some \$20,000 a year more than BScs (Table 2), and bachelors working in industry — in marketing and engineering — earn on average \$46,000 a year. Their academic colleagues of like age, who remain at the bench, earn a mean \$14,000.

A warning, though. People new to biotechnology are not necessarily novices. Though all of those in this subsample reported having less than 5 years' experience in biotechnology, one can make that shift at any age, at any level, from any industry, and from almost any background. □

Douglas McCormick, a freelance writer is at 145 Blauvelt Street, Teaneck, New Jersey 07666, USA. He was until recently editor of Bio/Technology magazine and editorial vice president of Nature America, Inc. John Hodgson is articles editor for Bio/Technology.

TABLE 2 — AVERAGE INCOME BY DEGREE LEVEL

Degree	Average experience in field	Average annual compensation
BSc	3.38	\$43,218
MSc	1.13	\$44,750
PhD	2.65	\$64,249

Respondents with less than five years industry experience

1. Spalding, B. J. *Bio/Technology* **11**, 768–770 (1993).
2. Burrill, G. S. & Lee, K. B. *Biotech 93: Accelerating Commercialization*. (Ernst & Young, San Francisco, 1992).
3. McCormick, D. K. & Goodstein, M. *Bio/Technology* **11**, 994–996 (1993).